

Tanzania Threshold Country Program: End Term Evaluation

Final Draft

James Hollyer
Leonard Wantchekon

February 2011

Acknowledgements

The authors wish to thank those who's participation in the evaluation process made this report possible. We would particularly like to thank officials from the Government of Tanzania and current and former staff of implementers (Kilimanjaro International, PACT-Tanzania, Women for Law and Development in Africa, Campaign for Good Governance, OPDAT, ICITAP, NYC DoI, and Crown Agents Consulting) for generously providing their time to discuss their experiences during the implementation of the Tanzania TCP. We would also like to thank Aaron Karnell of USAID for providing invaluable perspective and information on TCP. And we would also like to thank Franck Wiebe, John Molyneaux, Malik Chaka, and Sophia Sahaf at MCC for their guidance and help throughout the project and the staff of EcomResearch for their role in both quantitative and qualitative data collection.

Abbreviations

AWF	African Wildlife Foundation
BONGA	Building Organizational Networks for Good Governance and Advocacy
CCM	Chama Cha Mapinduzi
CGG	Campaign for Good Governance
CSO	Community Service Organization
Deloitte	Deloitte & Touche
DPP	Department of Public Prosecution
ESAAMLG	Eastern and Southern Africa Anti-Money Laundering Group
FIU	Financial Intelligence Unit
ICITAP	U.S. Department of Justice International Investigative Training Assistance Program
IRT	Item Response Theory
MCC	Millennium Challenge Corporation
MP	Member of Parliament
NOLA	National Organization for Legal Assistance
NYC DoI	New York City Department of Investigations
OPDAT	U.S. Department of Justice Office of Overseas Prosecutorial Development Assistance and Training
PACT	PACT-Tanzania
PCCB	Prevention and Combating of Corruption Bureau
PETS	Public Expenditure Tracking Surveys
PPRA	Public Procurement Regulatory Authority
STR	Suspicious Transaction Report
SUNY	State University of New York
TCP	Threshold Country Program
USAID	United States Agency for International Development
WiLDaf	Women for Law and Development in Africa
WMA	Wildlife Management Area

Contents

1	Executive Summary	1
2	Methodology	10
3	Impact Findings	11
3.1	Public Procurement Reform	11
3.1.1	Enhancing the Capacity of the Public Procurement Regulation Authority	12
3.1.2	MP Training	24
3.2	Rule of Law	26
3.2.1	Training of the Prevention and Combating of Corruption Bureau . . .	27
3.2.2	Training of the Department of Public Prosecution	39
3.2.3	Legal Aid	41
3.3	Financial Intelligence Unit	44
3.4	Nongovernmental Sector Monitoring Capacity	49
3.4.1	Investigative Journalist Training	49
3.4.2	Public Expenditure Tracking Surveys	66
3.4.3	Public Awareness	73
3.4.4	Wildlife Management Areas	76
4	Conclusions and Lessons Learned	79
4.1	Lessons Learned	80
A	Statistical Appendix	88
A.1	Public Procurement	88
A.2	Rule of Law	91
A.3	Journalist Training	105
A.4	Public Expenditure Tracking	120
B	PCCB Bureau Chief Questionnaire	124
C	Questionnaire for PETS Council Chairs	130

List of Figures

1	Average Change in Compliance Scores	94
2	Effect of Training on Conviction Rate	95
3	Effect of Training on Trial Rate	96
4	Effect of Training on DPP Referrals	97
5	Effect of Training on Case Uptake	98
6	Simulated Difference btwn Powers-trained and Others	99
A.1	Change in Compliance Scores Not Attributable to Auditing	100
A.2	Disaggregated Effect of Training on Conviction Rate	101
A.3	Disaggregated Effect of Training on DPP Referrals	102
A.4	Disaggregated Effect of Training on Trial Rate	103
A.5	Disaggregated Effect of Training on Case Uptake	104
A.6	Distribution of Articles per Journalist per Month	107
A.7	Distribution of the Change in the No. of Articles per Journalist per Month	108
A.8	Enumerator Responsiveness to Article Quality	111
A.9	Estimates of and 95 Percent Confidence Intervals Around Article Quality	112
A.7	Distribution of Propensity Scores from Genetic Matching Algorithm	121
A.8	Histogram of Propensity Scores from Genetic Matching Algorithm	122

List of Tables

1	Project Costs by Component and Implementer	2
2	PPRA Compliance Indicators	14
3	Change in Overall Compliance as a Function of Auditing	17
4	Value of Procurements, 2005/6 Audits	19
5	Quality Questionnaire	53
6	General Effect of Training	58
7	Differentiated Effect of Training by Type of Training	59
8	Effect of Powers Session	62
9	Estimated Effect of Training on Article Quality	63
10	Change in the Percent of Questioned Expenditures Associated with PETS Training	70
11	Change in Auditor Opinions Associated with PETS Training	70

A.1	Eigenvectors From Principal Components Analysis	89
A.2	List of Newspapers Covered in the Data	106
A.3	Effect of Lake Zone Session	117
A.4	List of Newspapers in Sample and Partisan Affiliation	118
A.5	Estimated Effect of Training on Article Quality – Trained Authors Only . . .	119
A.6	Change in the Percent of Questioned Expenditures Associated with PETS Training – Matched Data	123
A.7	Change in Auditor Opinions Associated with PETS Training – Matched Data	123
A.8	List of PETS Councils Surveyed	123

1 Executive Summary

The Tanzania Threshold Country Program (TCP) provided \$11 million in support of anti-corruption programs between May 2006 and September 2008. The TCP sought to improve Tanzania's performance on the MCC's selection indicators (notably that for corruption) - which had previously proved insufficient to secure Compact eligibility.¹ Tanzania did qualify for a Compact Program in 2006, but nonetheless sought to implement the TCP.

The Tanzania TCP had four principal goals:

1. The reduction of corruption in public procurement.
2. An improvement in the enforcement of the rule of law.
3. The enhancement of the non-governmental sector's ability to scrutinize corruption.
4. The passage of an anti-money laundering act and creation of a Financial Intelligence Unit (FIU).

A total of fourteen interventions were performed under the Tanzania TCP, involving four Ministries, Departments or Agencies (MDAs), as well as a variety of non-governmental organizations. The TCP was managed under the aegis of USAID. Participants included Kilimanjaro International, Deloitte & Touche, PACT-Tanzania, the Campaign for Good Governance, Women for Law and Development in Africa (WiLDaF), the National Organization for Legal Assistance (NOLA), the African Wildlife Fund (AWF), and the US Department of Justice's International Investigative Training Assistance Program (ICITAP) and the Office of Overseas Prosecutorial Development Assistance and Training (OPDAT). A complete list of implementers and costs – sorted by Program component – is included in Table 1.

In September of 2009, the MCC commissioned this report to assess the performance of the Tanzania TCP. The purpose of this report is to assess whether or not the interventions under the Tanzania TCP had the desired effect, whether these interventions were conducted as intended, and whether these interventions proved sustainable. In answering these questions, the report is intended to provide lessons learned for both the MCC and the Government of Tanzania. Where possible, the report relies on quantitative methods to identify the causal effect of interventions launched under the TCP. These quantitative results are supplemented with qualitative work based on the evaluators' interviews with key

¹<http://www.mcc.gov/pages/selection/indicators>

TCP participants; document reviews; and semi-structured interviews conducted by EcomResearch, a Dar es Salaam-based research firm. Data for the quantitative analyses are drawn from a variety of sources - primarily from the Government of Tanzania and from Program participants - and were collected by EcomResearch.

Table 1: Project Costs by Component and Implementer

Component	Implementer	Cost
Public Procurement	Kilimanjaro International	\$2,600,000
	SUNY	\$130,000
	Subtotal:	\$2,730,000
Rule of Law	ICITAP	\$913,910
	OPDAT	\$33,000
	Kilimanjaro International (DPP Training)	\$609,274
	Kilimanjaro International (NYC DoI)	\$75,400
	Crown Agents Consulting	\$146,166
	WILDaf	\$640,805
	Japan Africa Marketing (PCCB Vehicles)	\$375,946
	Subtotal:	\$2,794,501
Financial Intelligence Unit	Deloitte and Touche	\$1,194,685
	Subtotal:	\$1,194,685
Non-Governmental Sector	PACT-Tanzania	\$2,800,000
	Campaign for Good Governance	\$200,000
	National Organization for Legal Assistance	\$322,740
	African Wildlife Foundation	\$50,000
	Subtotal:	\$3,372,740
Other	DPK Consulting (Judiciary)	\$326,881
	US Energy Association	\$15,000
	Macro International (M&E)	\$150,000
	Technical Services and Other Procurements	\$455,123
	Subtotal:	\$947,004
		Total: \$11,038,930

Cost figures are 'Accrued Expenditures' as of September 30, 2008, as reported to the authors by USAID.

A summary of the findings for each activity follows:

1. **Improve the capacity of the Public Procurement Regulatory Authority (PPRA) and conduct 38 audits of procuring entities.** (\$2,600,000) Threshold funds supported the training activities for the PPRA and were used to conduct a series of 38 audits over the space of two years. Quantitative findings reveal that the auditing process substantially improved the compliance of procuring entities with the Public Procurement Act of 2004. Results suggest that most of this improvement was a result of auditing, not of a general improvement of procurement quality over time. One audit report also uncovered irregularities in the procurement of electrical generators by Tanesco - the national electricity purveyor. The costs of the Tanesco mismanagement were estimated by a Parliamentary investigation to exceed \$68 million. A resultant scandal led to the resignation of the prime minister, as well as of several other members of the government. While our findings document an effect from the auditing program; it is less likely that the training seminars also implemented under the TCP had the desired effect on the PPRA's capacity.
2. **Strengthen the capacity of the Prevention and Combating of Corruption Bureau to investigate corrupt activities.** (\$1,168,476)² Several training activities - provided by a variety of program participants - were conducted with threshold support. Training involved informational seminars provided by ICITAP, the New York City Department of Investigations (NYC DoI), OPDAT, and Crown Agency Consultants. Using data provided by the PCCB and ICITAP, we analyzed the quantitative effect of the ICITAP training sessions. Our analysis did not uncover any conclusive evidence indicating that training improved PCCB performance. The ICITAP, OPDAT, and Crown Agents training sessions were similar in time frame and training methodology - which raises doubts about the effectiveness of these other trainings. However, we lack sufficient data to adequately assess the effectiveness of training provided by bodies other than ICITAP. At the request of the PCCB, NYC DoI embedded several of its staff with the PCCB in two ongoing investigations. One of these investigations - involving the construction of the Bank of Tanzania twin towers - led to high profile prosecutions and the resignation of the Chairman of the Central Bank. Reports from program participants suggest that this embedding procedure also had an indirect ef-

²This amount includes funds provided to ICITAP, OPDAT, and Crown Agents Consulting for training service, and funds awarded to Kilimanjaro International for the logistical coordination of training services provided by NYC DoI.

fect on corruption by facilitating the transfer of skills from NYC DoI agents to PCCB officials.³

3. **Train incoming recruits to the Department of Public Prosecution (DPP) in anti-corruption law.** (\$609,274) A full incoming class of some 157 recruits received training in anti-corruption law. Eight training sessions of three days or one week were provided, covering topics such as criminal prosecution skills; writing legal opinions; cyber-crime, money-laundering and fraud; and fighting corruption. This training facilitated the ‘civilianization’ of criminal prosecutions, which granted substantially more authority to the DPP. The civilianization process began after the start of the TCP, and this training program was added to the TCP to provide support. It is not possible to quantify the effectiveness of this training. However, the Government of Tanzania funds similar training for current recruits to the DPP - which suggests that Tanzanian officials found the training beneficial.
4. **Improve the provision of legal aid services.** To that end, establish a **legal aid secretariat** and provide **training for legal aid clinics.** (\$640,805) Women for Law and Development in Africa (WiLDaF) established 5 model legal aid clinics under this intervention. WiLDaF also helped create a legal aid secretariat incorporating 8 of the 12 major legal aid providers in Tanzania. A lack of data prevents us from quantifying the effectiveness of this intervention. Reports collected in Macro International’s Performance Audit suggest that many members of the public did not believe that their complaints would result in either recourse or punishment of offenders.
5. **Pass an anti-money-laundering and financing of terrorism law and establish a Financial Intelligence Unit (FIU) to aid in the enforcement of this law.** (\$1,194,685) An anti-money-laundering bill was passed in late 2006. The FIU was created in September 2007. However, implementation suffered from numerous delays, primarily resulting from the need for Parliament to approve the selection of a Commissioner for the FIU. The contractor successfully procured office materials and equipment for the FIU and helped to arrange training for FIU staff. However, staffing delays and limitations continued for the duration of the TCP. The FIU also did not receive any suspicious transaction reports (STRs) before the close of the TCP. Given challenges it is unlikely that the FIU had any effect on money-laundering or corruption in these years. While the FIU has secured funding to hire more staff and has begun to collect

³Interviews with PCCB and DoI officials, as well as the Program Manager from USAID.

STRs; it remains to be seen whether this institution will have any effect on money-laundering and corruption in the future.

6. **Improve the ability of the media to detect and report corrupt activities. This entailed training investigative journalists and providing training to both journalists and editors on the content of the anti-corruption statutes.** (\$184,960)

This report employs a quantitative analysis to assess the effectiveness of the investigative journalist training in increasing both the quantity and quality of articles relating to corruption and good governance. On average, training *did not* lead to an increase in the quantity of articles produced. However, one training session, in which participants were given travel grants to develop stories, *did* produce a permanent increase in the rate with which journalists produced articles. On average, training appears to *reduce* the quality of articles produced by trained journalists. This average result reflects the negative effects of training on articles that appear in partisan affiliated newspapers and magazines. We surmise that investigative journalists use their training to produce one-sided attack articles in partisan publications. Training had a positive, but statistically insignificant, effect on the quality of non-partisan publications.

7. **Increase the use of Public Expenditure Tracking Surveys (PETS) to report on the budgeting activities of local governments.** (\$2,815,040)⁴ 77 district in Mainland Tanzania established PETS councils and received training from one of two implementers - PACT-Tanzania (PACT) and the Campaign for Good Governance (CGG). PETS councils examine the financial records of local governments to determine whether budget priorities are in keeping with decisions of village councils and to detect any financial mismanagement. Interviews with current PETS council chairs reveal that these councils continue to meet after the close of the Tanzania TCP and that most report having detected mismanagement. However, we are unable to detect quantitative evidence for the effectiveness of the PETS intervention. Between FY 2006/7 and FY 2008/9, districts trained in PETS techniques do not reduce levels of questioned expenditures, nor do they demonstrate an improvement in auditor ratings, relative to non-PETS districts. The absence of a quantitative effect may indicate that the episodes of mismanagement detected by PETS committees involved small sums, such that overall budgets were relatively unaffected despite the efforts of PETS councils.

⁴This sum includes funds awarded to both PACT-Tanzania (\$2,615,040) and to CGG (\$200,000).

Alternatively, it may suggest that the detection of mismanagement does not result in the punishment of government officials, nor in the rectification of government accounts.

8. **Increase public awareness of corruption and of the contents of the anti-corruption law.** (\$322,740) The National Organization for Legal Assistance (NOLA) conducted several public awareness activities, including: media campaigns, the creation of billboards and leaflets, public discussions regarding the Prevention and Combating of Corruption Act of 2004, and the training of journalists and editors on the particulars of this Act. While this intervention was fully implemented; participants expressed concern that the duration of the campaign was insufficient.⁵ Moreover, portions of the public awareness campaign were poorly drafted and had to be withdrawn. On the other hand, members of the media who received training on the Prevention and Combating of Corruption Act reported positive assessments of this training and expressed hope that further training might be continued.
9. **Reduce corruption in the wildlife and game sector by supporting the drafting of business plans for Wildlife Management Areas (WMAs).** (\$100,000) WMAs were created by the Government of Tanzania to devolve control over wildlife management to local-level organizations. Management of village lands adjacent to national parks was granted to local level community organizations. It was hoped that this would give villagers an incentive to sustainably manage wildlife in these lands and that community oversight might curb corruption in the wildlife and game sector. Prior to the TCP, the African Wildlife Foundation (AWF) had partnered with community organizations in three WMAs to help with these efforts. The TCP provided funds to the AWF to draft business plans to direct private sector investment in two of the WMAs in which it was involved. These plans were drafted and one of these WMAs successfully secured a private sector investor. The AWF reports that this investment dramatically increased village revenues from wildlife management - though we are unable to obtain sufficient evidence to corroborate this claim.

The Corruption Environment:

The TCP was implemented at a time of significant changes in the corruption environment in Tanzania - particularly in the political salience of corruption. The president of

⁵Macro International's audit noted its concerns that sustaining public awareness will be a 'challenge.' It was noted that campaigns may continue, depending on levels of interest.

Tanzania – Jakaya Kikwete – was elected in 2005 based largely on an anti-corruption campaign message.⁶ Prior to President Kikwete’s election, the passage of the Prevention and Combatting of Corruption Act of 2004 increased the range of prosecutable offenses under Tanzania’s legal code. During the course of the TCP, a number of high profile corruption investigations were launched – including investigations into a military contract awarded to BAE,⁷ the mishandling of accounts by the Central Bank, the procurement practices of the state electricity producer TANESCO, and into the construction of the Bank of Tanzania Towers in Dar es Salaam. Several of these investigations directly involved TCP support as is outlined in further detail below.

Over the period from 2005, Tanzania has seen a substantial improvement in its score on the World Governance Indicators’ Control of Corruption measure, published by the World Bank Institute. A marked improvement took place from 2005 to 2006, when Tanzania jumped from a score of -0.72 to -0.28.⁸ As noted above, this improvement was sufficient for Tanzania to qualify for a compact program before the TCP was implemented. Subsequently, Tanzania’s control of corruption score has declined slightly – though it continues to meet the criteria for a compact program.

While changes in the political salience of corruption likely bode well for Tanzania; they make the isolation of the effects of the TCP more challenging. It is difficult to isolate the role played by the TCP, as opposed to changes in the Government of Tanzania’s reaction to corruption that would have occurred absent the TCP. It is possible that the TCP played a role in increasing the political salience of corruption in Tanzania – that the Program played a role as a ‘scene setter.’ However, we are unable to conclusively document such an effect. For that reason, wherever possible, we rely on quantitative estimates that are able to isolate the effect of TCP interventions from overall changes in the corruption environment.

We should also note that the political environment in Tanzania is dominated by a single party – the Chama Cha Mapinduzi (CCM). Prior to 1992, Tanzania was a single party state led by the CCM (Mmuya and Chaligha, 1992, 1994). While reforms have ushered in multi-party democracy; the CCM has continued to hold the presidency and has an overwhelming majority of parliamentary seats (264 of 324). The powerful role of the CCM may have altered the effectiveness of several of the interventions carried out as part of the TCP.

⁶Muga, Emmanuel. “Can election help Tanzania tackle corruption.” *BBC News*. October 27, 2010.

⁷BAE (formally British Aerospace) is a UK-based defense and aerospace contractor.

⁸The World Governance Indicators scores range from -2.5 to 2.5, with higher scores indicating lower levels of corruption.

Lessons learned:

- **Embedded training exercises are more effective than short-term training seminars.** Much of the Tanzania TCP focused on training civil servants or members of civil society. Training generally took one of two forms: short term seminars (generally of one week or less), or more intensive training, often involving side-by-side work with embedded trainers. Quantitative and qualitative results suggest that many of the more intensive forms of training had positive effects. By contrast, we were unable to detect any effect of short training seminars offered. It must be cautioned, however, that we cannot claim that this association is causal. The forms that training assumed were dictated – at least in part – by the availability of staff for training and the willingness of participating MDAs to accept embedded trainers. In some instances, longer term or embedded training may not have been possible.
- **The two year timeframe may limit the effectiveness of some interventions.** The limited timeframe for the TCP may have had both benefits and drawbacks. Program participants note that the short implementation period may have increased the urgency with which implementation took place.⁹ However, the two year period proved insufficient for some interventions to be completed and may have hindered the effectiveness of others. Given the difficulties in assessing participants' sense of urgency and the absence of variation in time limitations, we cannot conclusively document the purported benefits of a short implementation period. The drawbacks are more readily visible: For instance, the creation of the FIU suffered from numerous delays and was not fully operational by the close of the TCP. Similarly, programs to promote public awareness of corruption may be expected to have less sustainable effects due to the limited period of implementation.
- **The TCP often achieved successes by contracting implementers with established relationships with Tanzanian government agencies.** This lesson is most clearly in evidence for the interventions targeting the PCCB and the WMAs. The PCCB specifically asked USAID to contract a portion of the Rule of Law component of the TCP to NYC DoI due to the long relationship between the two agencies. Similarly, USAID contracted the WMA intervention to the AWF – an organization long involved in the Burunge and Endiumet WMAs. Both interventions anecdotally benefited from the level of trust between implementers and other participants and from the knowledge

⁹For instance, in structured interviews with PCCB Bureau Chiefs.

implementers had of local realities. These two implementers were able to build on past successes in their interactions with Program participants and from close professional ties between implementers and participants. These results suggest the value of (1) contracting implementers with a proven track record of successful interactions with Program participants, (2) heeding the advice of Program participants in the selection of implementers, and (3) hiring multiple implementers rather than one prime contractor to implement all interventions.

- **The flexibility of USAID to new opportunities allowed for several well-targeted interventions to be implemented.** This finding holds particularly true for the intervention targeting the DPP and the WMAs. Both interventions were added to the TCP in response to changes in Tanzanian policies. USAID responded to these changes and was able to support several government programs as a result. In large part, the flexibility of the Tanzania TCP reflects the skilled efforts of USAID staff – particularly the program manager, Aaron Karnell. It is also probable that this program flexibility produced appropriate interventions due to the clear definition of the Tanzania TCP’s overall mission – namely the reduction in corruption.
- **Greater selectiveness was warranted to ensure that interventions targeted corruption.** Several of the interventions undertaken as part of the TCP bore a questionable – or, at best, an indirect – relation to the goal of combating corruption. The link between corruption and the expansion of the legal aid system was particularly tenuous. Some of the training sessions administered to the PPRA – for instance workshops in team building – similarly bear at best an indirect, or little, relation to combating corruption.
- **Interventions that sought to create new institutions often proved more sustainable than capacity-building exercises.** A variety of new institutions were created as part of the TCP. While we are unable to document an effect on corruption from several of these institutions; all remain in operation two years after the close of the TCP. This effect may not be causal – it may be that participants were more willing to engage in institution-building in areas that were more likely to produce sustainable effects. However, the construction of new institutions acts as a strong predictor for the sustainability of an intervention.
- **Greater attention to M & E from the inception of the Program would like improve both the ability to detect the impact of interventions and may have pre-**

vented the implementation of interventions without a clear logical connection to the aims of the Program. Assessment of the impact of the Tanzania TCP was frequently hindered by the limited availability of data and the difficulty of constructing empirical strategies to identify the effect of interventions after the fact. Were M&E more thoroughly integrated from the inception of the Program, it is likely that indicators could more closely mirror the intended effects of interventions. Problems of causal identification may have been eased through greater use of randomization of treatment (e.g., training) or in the timing of treatment. Subsequent evaluations might also have been better able to assess the effectiveness of the Program in producing its desired outcome – reductions in the level of corruption – rather than focusing on outputs – e.g., the behavior of anti-corruption agencies, or of trained journalists. Finally, a greater focus on M&E would have necessitated careful consideration of the logic of all interventions, and may have prevented the implementation of interventions whose logic was questionable.

2 Methodology

This *ex post* evaluation is intended to provide an independent analysis of the impact and effectiveness of the Tanzania TCP. It is intended to assess the impact of Program activities; to provide lessons learned for the Government of Tanzania, the MCC, and program participants; and to uphold the MCC's commitment to measuring program results. To achieve these ends, this evaluation does the following:

1. Wherever data and identification constraints allow, estimate the causal effect of the intervention with quantitative methods. Where this is not possible, we attempt to assess the effectiveness of the intervention with qualitative methods.
2. Assess whether Program activities were implemented as intended.
3. Assess whether the results of these activities will be sustainable.
4. Determine what lessons may be drawn from the successes or failures of Program activities.

The methodologies employed by the evaluation are both quantitative and qualitative. Wherever possible, we obtain data and employ an identification strategy with which to

estimate the causal effect of a given intervention. Data are collected from program participants, and data collection was handled by EcomResearch – a Dar es Salaam based research firm. Because data is often unavailable, and because randomization was never employed in treatment assignment, such causal estimations are sometimes impossible. Where quantitative methods cannot be used to assess the effect of an intervention, we employ qualitative methods. We conduct interviews with program participants, engage in document review, and – in some cases – conduct semi-structured field interviews to assess the implementation and sustainability of a given intervention. We also examine the logical link between the intervention and the reduction of corruption, in light of related academic literatures. The quantitative methods employed in this evaluation give precise estimates of the effect of training on a narrowly defined quantity of interest. Qualitative methods, on the other hand, cannot give as precise an estimate, but may incorporate a broader variety of factors than quantitative estimation allows.

Since the evaluation of each intervention differs slightly depending on data availability, we discuss the methodology for our assessments in more detail in the Impact Findings section below.

All estimates, however, are based on information made available in late 2009 and early 2010. The TCP closed in late 2008. The passage of time since the close of the TCP is unlikely to affect the quantitative estimates produced in this evaluation. However, it is likely that qualitative evidence – for instance evidence based on interviews – may be affected by the passage of time, as memories of precise details of implementation efforts may have faded.

3 Impact Findings

3.1 Public Procurement Reform

One of the primary goals of the Tanzania TCP was to reduce corruption in the public procurement process by improving the ability of the Public Procurement Regulatory Authority – the body charged with oversight authority over all government procurement in mainland Tanzania – to oversee procurement decisions. In addition, an intervention was also planned to train Tanzanian members of parliament (MPs) in public procurement law.

3.1.1 Enhancing the Capacity of the Public Procurement Regulation Authority

Background

Kilimanjaro International, the implementer, was awarded \$2.6 million to conduct audits of 40 procuring entities and assess their level of compliance with the Public Procurement Act of 2004 and subsequent PPRA regulations. It was also to provide training to PPRA staff and to arrange meetings between the PPRA board and US public procurement officials based in Washington, with the aim of disseminating best practices and building human capital. Finally, Kilimanjaro was to orchestrate eight meetings between local procurement stakeholders and PPRA staff. These meetings were to disseminate information regarding procurement regulations and to gather reactions and suggestions from procurement stakeholders regarding public procurement policy.

Methodology

It is difficult to adequately quantify the impact of most elements of the procurement component of the TCP. The effects of capacity building on the PPRA's performance, and of stakeholder fora are not readily amenable to quantitative analysis for two reasons: First, the outcome measure is non-obvious. For instance, it is difficult to isolate the performance of a trained PPRA staff member and compare her performance to an untrained member. Second, even if it were possible to isolate an outcome of interest, we do not have an adequate control group with whom to compare participants in the TCP. Although multiple rounds of training were provided to members of the PPRA (such that not all members were trained simultaneously); these rounds were held in rapid succession, such that there does not exist a period during which to compare those who were trained to those who were not. Changes in PPRA practices as a result of the Board's trip to Washington, DC were applied to the PPRA as a whole; there is no portion of the agency or country unaffected by these regulatory changes.

However, data in the audit reports – and in the follow-up audit reports¹⁰ – conducted by the PPRA and by Kilimanjaro permit a rough approximation of the causal effect of auditing.¹¹ We can also – under strong, but not unreasonable, assumptions – use these

¹⁰Audits were conducted on municipal, district and regional governments, as well as central ministries and state owned utilities.

¹¹We have received mixed reports as to the enforcement following auditing. Interviews with PPRA staff indicate that there exists a threat of sanction – including of removal – facing procurement staff who fail to respond to negative audit reports. However, Macro International's Final Performance Audit Report for the TCP notes that there was not a formal mechanism to enforce audit findings at the time the performance audit was drafted – in 2008.

reports to gain some quantitative measure of the cumulative effect of the other (non-auditing) elements of the procurement component of the TCP.

As our measure of compliance with procurement requirements, we rely on 13 indicators of compliance contained in the audit reports.¹² The PPRA, in cooperation with Kilimanjaro, constructed these compliance measures to assess changes in procuring entity compliance with procurement regulations over time.¹³ Audits conducted in 2007 focused on procurement activities during the 2005/6 fiscal year. The PPRA conducted follow-up audits of the same entities for fiscal year 2006/7.¹⁴ An additional set of audits and follow-ups were conducted for the 2006/7 and 2007/8 fiscal years, respectively.¹⁵

We attempt to estimate the effect of auditing on subsequent procuring entity compliance by comparing the initial audit scores to those published in the follow-up reports. We calculate the difference in reported compliance levels for 19 procuring entities between fiscal year 2006/7 and 2005/6, and the difference for 18 procuring entities between fiscal year 2007/8 and 2006/7.¹⁶ Initial audits were conducted jointly by Kilimanjaro and the PPRA; while follow-ups were conducted by the PPRA alone.

Some concern must be exercised in treating these results as an estimation of the causal effect of auditing. First, the initial and follow-up audits were conducted by different bodies. Though efforts were made to ensure that the compliance measures were comparable across audits; differences in the composition of auditing teams, in subjective judgments, and – however slight – in methodology may affect the validity of these comparisons. Second, simple comparisons across time do not adjust for factors other than auditing that may be driving changes in compliance. Even if changes in compliance are measured perfectly,

¹²Unfortunately, not all audit reports contain data on the dollar value of procurement. Therefore, we cannot assess the effect of auditing on procurement costs.

¹³These measures were not contained in Kilimanjaro's initial 2007 audit reports. Rather, the PPRA and Kilimanjaro constructed these measures based on the material in the initial reports. These 13 compliance measures have been published in all subsequent audit reports, including retroactive measures obtained from the initial Kilimanjaro audits.

¹⁴Follow-up audits were conducted by different entities than the initial audits. All follow-up audits were conducted by the PPRA in conjunction with external contractors other than Kilimanjaro. The follow-up audits were less thorough than the initial audits, and particularly concentrated on areas where the initial audits revealed compliance to be low.

¹⁵It should be noted that procuring entities may face punishment should they fail to improve on problem areas noted in the initial audits. The CEO of the PPRA noted, during an interview, that procurement staff face potential dismissal should attempts to address earlier failures prove to be inadequate (interview with PPRA chief executive officer, April 2010.) However, a performance audit conducted by Macro International found that the PPRA did not itself have direct power to enforce such punishments at the time these audits were conducted.

¹⁶The Ministry of Public Safety and Security, which was initially audited by Kilimanjaro, ceased to exist before follow-up audits were conducted.

we cannot attribute any results to auditing alone.

We attempt to control for trends in compliance behavior over time that may produce the latter form of bias. To do so, we compare the initial audit scores produced by the Kilimanjaro audits covering fiscal year 2005/6 to the subsequent PPRA initial audits of *different procuring entities* covering fiscal year 2006/7. If the two sets of procuring entities covered by these audits are sufficiently similar, any difference in average compliance scores may be attributable to trends over time.¹⁷ These differences may be induced by TCP activities aimed at strengthening the PPRA and at increasing the capacity of MPs to oversee procurement regulations. Though, other factors may be responsible for such trends as well. The estimated causal effect of auditing will, therefore, be the difference in compliance scores in the initial fiscal year 2005/6 audits and the subsequent fiscal year 2006/7 audits, minus the difference between compliance scores in the initial 2005/6 audits and the set of 2006/7 audits performed on different procuring entities.

The thirteen compliance indicators used in the audit reports are enumerated in Table 2.¹⁸

Table 2: PPRA Compliance Indicators

Tender Board	Existence of a tender board in compliance with the Public Procurement Act of 2004 and with PPRA regulations.
PMU	Existence of Procurement Management Unit (PMU) in compliance with Public Procurement Act of 2004 and with PPRA regulations.
Interference	Percentage of tenders without interference between the audit office, tender board and PMU.
Procurement Plan	Preparation of an annual procurement plan.
Compulsory Approvals	Percentage of tenders which received compulsory approvals.
Pct. Advertised	Percentage of tenders for which bidding opportunities were advertised.
Pct. Publicized	Percentage of approved tenders that were reported to the public.
Pct. Compliance Time	Percentage of tenders where the time allocated for bidding was in compliance with the Public Procurement Act and PPRA regulations.
Pct. Auth. Method	Percentage of tenders in compliance with authorized procurement methods.
Pct. Stnd. Doc.	Percentage of tenders using standard tender documents.
Pct. Complete Records	Percentage of tenders for which the audit was able to discover complete records.
Quality Oversight	Existence of an inspection committee and project manager to ensure the quality of procured goods/services.
Contract Imp.	Percentage of contracts which have been implemented.

We additionally conduct a qualitative evaluation of the two non-auditing elements of this intervention – strengthening the PPRA, and conducting stakeholder fora. Since the

¹⁷The PPRA states that audits are conducted on a representative sample of procuring entities each year.

¹⁸All indicators are measured on a scale from 0 to 100.

quantitative analysis we conduct can only give a rough estimate of the effects of these two elements of the procurement component of the TCP on procuring entity compliance – and can we can only obtain this rough estimate for the year 2007 – we supplement our quantitative analysis with a qualitative analysis. The analysis relies on information garnered from interviews with program participants and program documents – quarterly reports and M&E audits. We attempt to speak to two questions: (1) Were there any problems in the implementation of training or stakeholder fora? (2) Given any issues in implementation, what is the likely link between these interventions and corruption in procurement.

Outputs

Kilimanjaro met all intended outputs.

Outcomes

Quantitative Results: PPRA Audits and Procuring Entity Compliance

To assess the change in the average level of compliance before and after the initial audits were conducted, we rely on a simple difference-in-means test comparing the levels of compliance found in the follow-up audits to those in the initial audits. We conduct this test to examine both the difference in compliance levels between the fiscal year 2005/6 audits and the 2006/7 follow-ups, and the difference in compliance levels for the full set of audits conducted by the PPRA (which also include 2006/7 initial and 2007/8 follow-up audits). Positive values indicate and improvement in the level of compliance following the initial audit, negative values indicate the reverse. Our estimates recover both the average difference in each compliance indicator and a 95 percent confidence interval around these estimates.

Figure 1 plots the average difference in compliance scores (black dots) and 95 percent confidence intervals (lines) for the full set of audits conducted by the PPRA. All estimated differences are positive, indicating an improvement in compliance levels. And most are significant at the 95 percent level. The lone exception is the percentage of tenders using standardized documentation – improvement in this measure is not statistically significant.

Similar results are obtained if we compare the differences in scores only for audits conducted in 2005/6 to follow-ups conducted in 2006/7. The details of this analysis are reported in Appendix A.1.

Compliance appears to improve most when initial levels of compliance are low. For instance, the existence of a procurement plan, the percentage of contracts publicized, and the percentage of tenders completed within statutory mandated time all witnessed large and significant improvements. This is partially because initial levels of compliance along these measures were quite low. On a scale from 0 to 100, the average score for the existence of a procurement plan was a 30.5 in the initial round; only an average of 28.2 percent of tenders required compulsory approvals; an average of 15.5 percent of contracts were publicized, and only an average of 30.1 percent of tenders were completed in the mandatory statutory time. Measures in which procuring entities initially scored more highly – such as the Tender Board and PMU measures – do not document as large an improvement.

This result suggests that there may be diminishing marginal returns to auditing.¹⁹ This suggests that the benefits to auditing may be diminishing over time, and that there is an optimal frequency with which audits should be conducted. It is also consistent with many models of accountability, wherein monitoring is necessary to ensure a minimal level of compliance. However, monitoring and accountability cannot push compliance beyond a certain level (see for instance Besley, 2006).

Unfortunately, because each entity only went through one round of auditing (including both initial audits and follow-ups), we cannot discern this optimal frequency in this report. Nor is it likely to be possible to establish the optimal frequency of auditing in from a Program with a two year time frame. To establish the optimal frequency of auditing, one would ideally conduct an initial audit of a large number of entities. One could then randomly assign the timing of subsequent audits to these entities. One could then empirically assess the effects of diminishing marginal returns from auditing and assess these against the costs of frequent auditing. Such a research design would require a lengthy time frame – potentially spanning 5 or more years.

In addition to comparing the 13 individual compliance indicators, we compare changes in a measure of overall levels of compliance between the initial round of auditing and the follow-ups. While the PPRA constructs a measures of overall compliance by taking the simple average of the 13 compliance indicators; we prefer not to rely on this measure since it implicitly gives each of the 13 compliance measures equal weight. Instead, we perform a principal components analysis of the 13 compliance indicators (Kim and Mueller, 1978). Details on the construction of this measure are provided in Appendix A.1.

¹⁹Though it may also be attributable to the fact that compliance measures are truncated – no entity can improve beyond a score of 100.

The index of overall compliance we construct has a mean value of zero and a standard deviation of 2.16. Higher values denote improved performance. To determine whether an improvement in overall compliance levels took place between the initial and follow-up audits, we regress the value of this index on an indicator for follow-up audits. This indicator takes the value one if the audit is a follow-up audit and the value zero if it is an initial audit. The value of coefficient on this indicator variable will thus denote the average difference in initial and follow-up overall compliance scores. We perform this regression both on the full set of audits and on those audits covering the 2005/6 and 2006/7 fiscal years. Results are reported in Table 3.

Table 3: Change in Overall Compliance as a Function of Auditing

	Full Sample	FY 2005/6-FY 2006/7
follow-up	3.084*** [2.363,3.805]	3.295*** [2.095,4.495]
constant	-1.542*** [-2.052,-1.032]	-1.778*** [-2.638,-0.917]
$\hat{\sigma}^2$	1.512	1.745

Results of a regression of the overall compliance index on an indicator for follow-up audits. The coefficient on the follow-up variable represents the average difference between initial and follow-up levels of compliance. A positive coefficient represents an improvement, a negative the reverse. 95 percent confidence intervals are reported in brackets below the coefficient estimates. * represents significance at the 95 percent level. ** represents significance at the 99 percent level. And *** represents significance at the 99.9 percent level. $\hat{\sigma}^2$ denotes the standard error of the regression – roughly interpreted as the average difference between the observed outcome and the prediction of the regression.

The coefficient on the follow-up term is substantively large and highly significant, in both samples. On average, there is a roughly 1.5 standard deviation increase in overall compliance levels between the initial audit and the follow-up audit.

The results displayed above capture the difference between initial and follow-up audits conducted in two different fiscal years. It must be emphasized that these results do *not capture the causal effect of auditing*. Changes in any factors that may affect compliance

with procurement regulations during the year between the initial and follow-up audits may drive the results. This possibility is particularly significant for assessing the difference between the initial fiscal year 2005/6 audits funded by the MCC and the fiscal year 2006/7 follow-ups to these audits conducted by the PPRA. Since the MCC and the PPRA were cooperating in other activities that may affect levels of procuring entity compliance – e.g., the training of PPRA staff – during the 2007-2008 time period, it may be that these other activities, rather than auditing, drive the observed difference in compliance rates over time. Compliance rates may also be improving for reasons exogenous to the TCP.

It is possible, however, to estimate the effect of auditing as opposed to the effect of other factors. To do so, we compare the difference in the compliance levels recorded in the initial fiscal year 2005/6 audits to the compliance levels in the initial fiscal year 2006/7 audits. The difference between this comparison and those comparisons reported above is as follows: In this set of estimates, we do not compare the initial audits to the follow-up audits performed on the same set of procuring entities. Rather, we compare the initial audits on *one set* of procuring entities and compare these results to the initial audits on *another set* of procuring entities taken at a later point in time.

If there exist factors other than auditing that change compliance with procurement regulations (e.g., MCC training, changes in procurement regulations/best practices, etc.), then these should affect the compliance levels of the comparison group.²⁰ If, however, procuring entities simply respond to auditing by improving compliance levels in areas deemed lacking by the auditors, then we should only witness changes in compliance levels between initial and follow-up audits. By comparing the difference between the initial audits conducted in 2005/6 and 2006/7, to the difference between the initial and follow-up audits, we can gain some sense of the relative effects of auditing and of other extraneous factors.²¹

We report the results of this analysis in A.1. Figure A.1 indicates that none of the indicators of compliance exhibit significant trends over time. We therefore conclude that the improvements in compliance scores following auditing are the result of procuring entity

²⁰Thus, we should observe differences between compliance levels in initial fiscal year 2005/6 audits and the initial fiscal year 2006/7 audits.

²¹A critical assumption underlying this analysis is that the procuring entities selected for auditing in 2007 are similar to those entities selected for auditing the following year. Otherwise differences in initial compliance levels may simply be due to the differing nature of the entities being audited. The PPRA reports that procuring entities are selected for auditing each year such that a representative sample of procuring entities are audited. Entities are selected without replacement – such that a procuring entity that is selected for auditing one year will not face another audit until all other entities have been audited. (Interview with PPRA chief executive officer, April 2010.)

responses to audit reports rather than trends of improving compliance over time.

Table 4: Value of Procurements, 2005/6 Audits

Procuring Entity	Total Procurements, TzSh	Total Procurements, US Dollars
Arusha Municipal Council	929,889,750	\$797,839
Arusha Regional Council	101,712,381	\$87,269
Dar es Salaam Regional Admin. Secretariat	160,610,712	\$137,803
Dar es Salaam Water and Sewage Auth.	127,446,195,549	\$109,348,007
Dodoma District Council	518,119,175	\$444,543
Ilala Municipal Council	2,528,795,968	\$2,169,690
Iringa Municipal Council	1,261,129,864	\$1,082,041
Min. of Agric., Food Security and Cooperatives	32,555,590,585	\$27,932,485
Mbeya District Council	835,262,831	\$716,650
Min. Health and Social Welfare	207,575,248,050	\$178,098,213
Min. Public Safety and Security	8,194,720,688	\$7,031,017
Min. Water	102,046,765,000	\$87,555,461
Mwanza City Council	missing	missing
National Housing Corp.	1,203,755,370	\$1,032,814
Tanzania Airport Authority	4,036,103,472	\$3,462,951
TANESCO	missing	missing
TANROADS	100,006,000,000	\$85,804,498
Vocational Ed. and Training Author.	1,414,720,050	\$1,213,821
Min. Energy and Minerals	1,416,558,151	\$1,215,398
Min. Education and Vocational Training	missing	missing

Calculations for dollar values are based on the IMF's end of year exchange rate for 2005 of 1165.51 TzSh/Dollar. All additional calculations involving exchange rates are based on the IMF's end of year 2005 data.

While our results point to a change in the degree of compliance with procurement regulations resulting from auditing; they do not give an indication of the effect of compliance on procurement costs. To the extent that procurement regulations deter corruption, we would expect improved compliance to lower costs. Unfortunately, we lack adequate data to test for such an effect. Our records of the 2005/6 initial audits contain the value of procurements for nearly all audited entities, but the follow-up audits do not contain this information.

To give some sense of the potential economic magnitude of the effect of auditing, we document each procuring entities procurement costs in Table 4. The table lists total procurement costs in both Tanzanian shillings and US dollars.²² In some instances, values are missing because the auditors were either unable to obtain these figures, or because they doubted the veracity of those figures obtained. In 2007, district level governments an average of 2.8 percent of expenditures were deemed 'questioned' by external auditors

²²We employ the IMF's end of year 2005 exchange rate to make this conversion.

– 5 percent of districts had 6 percent of expenditures or more classified as questioned.²³ If similar numbers can be extended to the other procuring entities listed above, the potential gains from auditing far outweigh the costs of this activity – which were \$2.6 million for all Kilimanjaro’s efforts and we would estimate not more than \$2.3 million for the audits.

We additionally test to see if the effects of auditing systematically differ depending on the level of procurement. If compliance systematically improves in entities with small procurement volumes, but does not improve in entities with high procurement volumes, the economic effect of auditing will be quite limited. To assess whether this is the case, we regress the change in each compliance indicator on the level of procurement in FY 2005/6. In these results (unreported), the coefficient on the level of procurement is never significant and varies in sign across regressions. Nor, in separate regressions, do we find an association between initial levels of compliance and procurement amounts. These results suggest that auditing is equally effective in large and small procuring entities.

Qualitative Results

In this portion of the analysis, we rely on information provided during interviews with TCP participants and on materials – quarterly reports and M&E audits – composed during the course of the TCP. We attempt to delineate the logical link between these interventions and the outcome of interest – corruption in procurement. And we document likely threats that might prevent this link from translating into a positive effect. We also discuss issues regarding the implementation of these interventions, which may aid or hinder the realization of a causal effect on corruption.

Strengthening the PPRA

To strengthen the PPRA’s capacity, Kilimanjaro International arranged two activities: a trip by the PPRA Board to Washington, DC, and five training sessions with PPRA staff. The logic of the interventions holds that better trained regulators will be better able to police corruption. There is substantial reason to believe that this is true. Rauch and Evans (2000) find, in a panel of low and middle income countries, that the meritocratic recruitment of skilled bureaucrats substantially reduces levels of corruption. Lewis (2008) finds that skilled managers with stronger educational backgrounds improve their bureau’s performance in the US. Moreover, officials at the PPRA report that skill transfers from

²³ Controller and Auditor General’s Reports.

Kilimanjaro were among the most important benefits from the TCP.²⁴

The trip to Washington, DC was reported as particularly significant by members of the PPRA. Members of the PPRA Board found their meetings with US procurement officials highly informative. And a number of changes to procurement regulations followed this trip.

The training of PPRA officials seems rather less likely to have had an effect on corruption. The training sessions consisted of a set of five approximately three-day seminars and covered such topics as teamwork and communications.²⁵ It is improbable that the material covered in so short a time would substantially affect the ability of officials to carry out their regulatory tasks. Moreover, the topics covered seem rather tangentially related to the outcome of interest. An interim audit conducted by Macro International also documents problems in implementation of the training program – namely the limited availability of PPRA staff.

Based on subject interviews, we conclude that the most effective form of training provided to the PPRA staff was that obtained from partnering with Kilimanjaro in the auditing process. PPRA staff participated in the auditing process alongside Kilimanjaro International employees.²⁶ Kilimanjaro International and the PPRA also jointly developed the compliance indicators used in the audit reports, which are designed to comply with OCED-advocated methodologies.²⁷ These activities reflect the transfer of skills from Kilimanjaro to PPRA staff.

Stakeholder Fora

Kilimanjaro coordinated a series of fora between the PPRA and stakeholders in the procurement process. The intent of stakeholder fora was to (1) improve communication between stakeholders and the PPRA, and (2) solicit stakeholder input on procurement regulations.

There are reasons to suspect that such a program may be successful in these aims. It is certainly important that stakeholders are informed of the content of procurement regulations. However, much depends on the very specific details of participation in the fora. For the fora to be an effective means of both disseminating information on existing

²⁴Interview with PPRA officials, November 2009. Similar statements from PPRA staff are provided in Macro International's Final Performance Audit.

²⁵Interview with PPRA officials, November 2009.

²⁶Interview with PPRA officials, November 2009

²⁷Interviews with PPRA officials, November 2009; USAID Program Manager, April 2010; and PPRA chief executive officer, April 2010.

regulations and of soliciting input on regulatory changes, participation in these fora must be broad and representative. If only a small portion of stakeholders attend the fora, then such meetings will be an ineffective means of communicating the content of procurement regulations. Moreover, the input provided by these stakeholders may be biased if the set of participants is not fully representative. For instance, if only large contractors attend the fora, they will no doubt advocate changes to procurement regulations in their own interest. If small contractors are not present to advance contrary opinions, the PPRA would receive biased input.

Moreover, the information provided by such fora may depend strongly on the procedures under which they were held. For instance, Humphreys, Masters and Sandbu (2006) find that – in open fora with assigned ‘discussion leaders’ – forum participants tend to adopt the policy positions advocated by these leaders.

Based on the content of interviews with TCP participants and on materials produced during the course of the TCP, there is mixed evidence for the efficacy of these fora. On the positive side, the PPRA reports that a number of changes in existing procurement regulations followed these meetings. The PPRA points to these fora as instrumental to the amendment of the Public Procurement Act and the passage of three new sets of regulations.²⁸

However, the Macro mid-term audit report noted several implementation problems. First, Kilimanjaro International faced difficulties in arranging these fora due to the limited availability of PPRA staff. Second, the participation of stakeholders was reported as being low.²⁹ The final Macro audit notes that Kilimanjaro took steps to address these problems. For instance, holding zonal meetings, rather than national meetings in Dar es Salaam, increased stakeholder participation. Nonetheless, grounds for concern remain.

Sustainability

The auditing program supported by the TCP has been expanded to cover a larger number of procuring entities. Prior to the TCP, such auditing was not a regularized. Presently, such audits are conducted by the PPRA annually. Summary reports of the audits are released to the public; though the full audit results are not.³⁰ PPRA Board approval is required for the release of audit data, which may delay publication. Although Board approval was also

²⁸Interview with PPRA officials, November 2009.

²⁹While Macro noted problems of attendance; the level of attendance was not quantified.

³⁰Audit reports are available at http://www.ppra.go.tz/index.php?option=com_wrapper&view=wrapper&Itemid=136.

required during the TCP; project staff report that the TCP agreement required the timely release of audit information. It therefore seems that some of the gains made during the TCP have been sustained, but that the timeline for reporting audit findings may be less effective.

The best practices implemented after the meetings with US procurement officials and changes in procurement practices resulting from stakeholder consultations have been written into regulatory codes. The training sessions were simply too short to have a sustained effect on participants' behavior.³¹ And much of the content covered was too tangentially related to the daily tasks of the PPRA to be reinforced by daily practice.

Also, the PPRA held another of consultations with the public in addition to those sponsored by the TCP.³²

Summary of Results

- A quantitative analysis of the results of initial and follow-up audits reveals a substantial and significant improvement – a 1.5 standard deviation increase in an overall compliance measure – in procuring entity compliance with procurement regulations following auditing.
- The method by which the PPRA was partnered with outside experts from Kilimanjaro International during the auditing process likely facilitated skills transfers to PPRA personnel.
- Capacity building and stakeholder fora did lead to a number of changes in procurement regulations and to the amendment of the Public Procurement Act of 2004. However, both suffered from low levels of participation and the short duration of training sessions reduced the effectiveness of these activities. However, we are unable to measure the degree to which these problems hindered the effectiveness of capacity building interventions.
- Data limitations prevent us from directly assessing the dollar value of the benefits stemming from the auditing conducted by Kilimanjaro International and the PPRA. However, given the size of the procurement budgets of the agencies audited, even

³¹We do not have any information as to what extent training sessions continue and whether they mimic the format of those offered by Kilimanjaro International.

³²Interview with PPRA officials, November 2009. It was unclear, however, whether these consultations were continued through the present.

a small percentage point reduction in levels of improper expenditures would yield a reduction of corruption levels that exceeded the cost of this intervention. As audited procuring entities substantially improve their compliance with procurement regulations, one may reasonably expect that corrupt behavior declined sufficiently to justify the cost of this intervention.

3.1.2 MP Training

Background

SUNY, the implementer, received a contract modification to their existing USAID contract amounting to \$130,000 to provide training to MPs regarding the content of procurement regulations and the role of Parliament in monitoring procurement and corruption. SUNY was to train 100 MPs and some 30 parliamentary staff over the course of the TCP.³³ Training was to include a primer on the Public Procurement Act of 2004, an introduction to both Tanzanian and international best practices in procurement, and an exposition of the Tanzanian parliament's role under the Public Procurement Act and its relationship with the National Audit Office (NAO) and the PPRA.

However, this intervention was not renewed after the first year of the Tanzania TCP due to the Program Manager's concerns with the content of training and with coordinating SUNY's activities with the PPRA and Kilimanjaro.³⁴

Methodology

We conduct a qualitative evaluation of SUNY's MP training. This section evaluates the purpose of the MP training program and its logical link to the goal of reducing corruption in Tanzania.

We do not attempt to quantify the effect of training. Data are not readily available by which to compare the actions of trained MPs to those who were not trained.

Outputs

- 37 MPs were briefed in the contents of the proposed Anti-Money Laundering Law
- 330 MPs were briefed in the contents of the proposed Public Procurement reform law³⁵

³³Numbers drawn from SUNY's statement of work.

³⁴Interview with Program Manager, USAID. April, 2010.

³⁵Numbers drawn from TCP Quarterly Reports.

- Training was extended to 70 local councilors and tender board members from the Mwanza region³⁶

Outcomes and Sustainability

SUNY advanced the following logic for the training of MPs as a means to combating corruption: The Parliament must pass all future procurement-related legislation. Also, the Standing Committees of Parliament are charged with oversight of government ministries, departments and agencies responsible for procurement. Better informed MPs might, therefore, introduce stronger procurement legislation in the future, reducing opportunities for corruption. Trained MPs might also be able to more actively exercise scrutiny over the expenditures of the executive, diminishing corruption directly.

There are two likely reasons this logic may not hold: First, as noted in the SUNY Statement of Work, most legislation in Tanzania originates in the executive. The prime minister exercises *de facto* control over the drafting of legislation – and is selected by the President to oversee the executive’s politics in the parliament. While a well-informed MP may modify proposed legislation for the better, or may encourage the executive to introduce new procurement-related legislation; it is unlikely that legislators will introduce new legislation without prompting from the executive. Though, parliamentary procedures do *de jure* allow members to introduce private bills.

There are also reasons to doubt that MP training will have a direct effect as a result of Parliament’s oversight role. Oversight of procurement activities is often highly technical and has distributive implications for concentrated interests. A substantial theoretical literature in both political science and economics suggests that such oversight activities are better carried out by bureaucratic than by elected officials (see for instance Alesina and Tabellini, 2007; Besley and Coate, 2003; Maskin and Tirole, 2004). Practically, it is unlikely that even trained MPs will uncover instances of procurement malfeasance on their own. However, trained MPs may conduct more thorough and sweeping hearings after initial evidence of corruption is uncovered by executive agencies or parliamentary staff.³⁷

The likelihood that MP training will affect parliamentary oversight also depends, in part, on the interests of the majority party. At the time of the TCP, the CCM party held 264

³⁶Information provided by USAID commenter

³⁷High quality legislative hearings may have an effect on the penalties for high level corruption. For instance, the legislative hearings following the Richmond affair helped to lead to the resignation of the then Prime Minister and several other cabinet ministers. Many of these legislators had received training by SUNY under a preexisting USAID sponsored program. However, the Richmond scandal was uncovered by the audits conducted by Kilimanjaro and the PPRA, not by MPs.

of 324 seats in the Union Parliament. The CCM also held the presidency. A lengthy literature in political science finds that legislative oversight of executive agencies tends to be more strict under divided government, at least in the United States (e.g. Mayhew, 1991). Huber and Shipan (2002) find that the professionalism of members of the legislature increases oversight of the bureaucracy, but that this effect is far greater under divided than under unified government. These findings do not suggest that increased MP knowledge of procurement policy will not have an effect on corruption, but indicate that any effect might be smaller than would be expected under divided government.

The above caveats suggest that even were MP training perfectly implemented, the effect of this training on corruption in procurement was likely to be marginal. However, as mentioned above, the Program Manager noted several problems with the implementation of the MP training. Both the caveats with respect to the logic of MP training and the problems with implementation reduce the prospects for positive outcomes from and the sustainability of this intervention.

Summary of Results

- Both the content and coordination of MP training were deemed unsatisfactory by the Program Manager.
- Both the academic literature and political context in Tanzania suggests that the prospects for the success of this intervention were low, even absent any implementation difficulties.

3.2 Rule of Law

The TCP aimed to improve the observation of the rule of law and the practice of good governance in Tanzania. To this end, the following five interventions were planned:

1. training the Commission for Human Rights staff in alternative dispute resolution and corruption investigations,
2. training public prosecutors and magistrates in anti-corruption legislation,
3. training investigators in anti-corruption investigations,
4. strengthening the capacity of the Institute for Judicial Administration and the legal faculties in Tanzanian universities to teach anti-corruption laws,

5. the establishment of a legal aid secretariat.

During implementation, several modifications were made to the program. These included the removal of CHRAGG training – as it was not seen as in keeping with the general focus on anti-corruption activities. Capacity building exercises for the legal faculties at Tanzanian universities or the Institute for Judicial Administration were also canceled. One new intervention was introduced; training for a new class of recruits for the Department of Public Prosecution (DPP) was added to the TCP.

3.2.1 Training of the Prevention and Combating of Corruption Bureau

The Prevention and Combating of Corruption Bureau (PCCB) is the primary law enforcement agency charged with corruption investigations and, as such, it received training in anti-corruption investigation methods. Four implementers provided this training: the International Criminal Investigative Training Assistance Program (ICITAP) and the Office of Overseas Prosecutorial Development Assistance and Training (OPDAT), both of the US Department of Justice; the New York City Department of Investigations (NYC DoI); and Crown Agency Consulting.

ICITAP Training

Background

ICITAP – a branch of the Department of Justice – provides technical assistance to law enforcement agencies worldwide. It received a \$913,000 project budget. ICITAP was previously involved in efforts to train the police force in criminal investigation techniques. In addition to training in investigative methods, ICITAP was to provide training for a criminal records management system for the PCCB. It also provided training in internal affairs to help detect and deter corruption within the PCCB.

In addition to training, ICITAP provided technical advice to the PCCB. While funds were insufficient to provide a full-time technical advisor to the PCCB; an ICITAP consultant acted as an intermediate technical advisor.

Methodology

To determine whether or not ICITAP training improved the PCCB's ability to conduct corruption investigations, we compare the PCCB's performance in regions that participated

more heavily in training to its performance in regions that participated less heavily in training. To do this, we rely on data provided by the PCCB listing the number of referrals received, number of new cases launched, the number of cases referred to the DPP and other prosecutors, and the number of convictions secured in each of 24 regional districts in mainland Tanzania. ICITAP has provided us with lists of program participants and the district offices from which they were drawn. We test whether performance in a given region depended on the percentage of district offices sending officials to receive training.

Improvement in the quality of policing – through efforts such as training – may be expected to have theoretically ambiguous effects. On the one hand, better quality policing may serve to deter corruption from ever taking place (Becker and Stigler, 1974; Tsebelis, 1989). If this is true, the number of instances of corruption – and perhaps the number of investigations, prosecutions, and convictions relating to corruption – would be expected to decline following training. On the other hand, the greater technical capabilities of the police force may encourage investigators to more actively pursue and prosecute cases. Recent evidence from Alt and Lassen (2010) suggests that greater prosecutorial resources increased the number of corruption convictions across US states.

In this instance, we do not expect training to have an immediate deterrent effect on corruption. If the training of the PCCB deterred corruption, it must be the case that those engaged in corruption were aware that this training was taking place. It seems unlikely that such individuals could be fully informed as to the existence of training and as to which district offices sent officials to receive such training. Instead, any deterrent effect would only manifest itself *after* an initial increase in the number of successful investigations or prosecutions. After individuals became aware of any increase in the PCCB's capabilities, they would adjust their behavior accordingly.

In light of this assumption, one would expect that – all else equal – increasing the technical capabilities of the PCCB would lead to a higher number of convictions as a percentage of prosecutions. However, this need not be the case. Improvements in the quality of investigations may make prosecutors more willing to bring difficult cases – that previously would be dropped – to court (see for instance Gordon, 2009). Or it may make the PCCB more willing to take up referrals that previously would go un-investigated due to the difficulty of making a case. But note what this implies: If training had an effect, *either* the percentage of convictions relative to prosecutions should rise *or* the percentage of cases brought to prosecution or percentage of referrals investigated should rise.

To assess whether either of these patterns resulted from ICITAP's training of the PCCB,

we employ a difference in differences specification.³⁸ We run a series of regressions of the following form:

$$\Delta y_{i,t} = \beta_0 + \beta_1 \Delta fraction\ trained_{i,t-a} + \beta_2 \Delta fraction\ trained_{i,t-a-1} + \dots + \beta_{b+1} \Delta fraction\ trained_{i,t-a-b} + \epsilon_{i,t}$$

where i represents region i , t represents month t , a and b specify the number of lags to include in the regression, and Δ is the first-difference operator. $\epsilon \sim N(0, \Sigma)$, where Σ is specified to allow each region's disturbance to be correlated over time, and $E[\epsilon_{i,t} | \Delta X_{i,t-a} \dots \Delta X_{i,t-a-b}] = 0$.³⁹

Our regressor $\Delta fraction\ trained_{i,t}$ represents the fraction of districts trained in a given region in a given month.⁴⁰ Thus, this variable assumes values between zero and one, and only assumes non-zero values in months where training takes place. Training is only expected to have an effect with a lag – it takes time for participants in training to confer their newfound knowledge to other members of their office, and there is an inherent lag from the point when this knowledge can be employed in investigations to when it will affect prosecutions. We base our model specifications on discussions with members of the PCCB on the amount of time necessary to investigate and prosecute a case, and always include a window of several months for an effect to be made manifest (e.g., between 1 and 4 months for the ratio of new cases to referrals, between 6 and 12 months on the ratio of convictions to prosecutions). These specifications allow for the effect of training to manifest itself over time and preserve the benefits of the difference-in-differences specification in causal identification (this estimation method is similar to that in Laporte and Windmeijer, 2005).⁴¹

The variable $\Delta fraction\ trained$ is defined so as to incorporate all forms of training

³⁸By employing this specification we control for time-invariant factors that may drive differences in our measures across regions. This specification is particularly vital given that the selection of officials for training was non-random.

³⁹Given that the data are observed monthly, the danger of autocorrelation of the standard errors may also be present. First differencing both sides of the regression equation will eliminate the danger for first-order autocorrelation. However, there remains a danger of higher order autocorrelation – particularly in time series data with frequent observations. We test for such autocorrelation by plotting the distribution of the residuals against time, and find no evidence for serial correlation of the standard errors.

⁴⁰Alternative specifications using the raw number of officials from each district trained in a given month as a regressor yield similar results.

⁴¹We have also employed specifications that use a cubic polynomial of time from training to identify the changing effects of training over time. Alternative numbers of lags to those reported below have also been employed. And we have employed models controlling for general time trends. Results are substantively unchanged to those reported below.

that were targeted at the district level. We do not estimate the effects of training targeted at members of the PCCB headquarters, as we do not possess measures of the regressand that are specific to PCCB headquarters.⁴² The training of members of the headquarters staff is thus assumed to affect all districts equally. Training sessions included seminars in Advanced and Basic Economics and Financial Crimes, Intermediate Fraud and Financial Crimes, Instructor Development, Internal Affairs Development, and Supervision and Management.

Finally, we supplement our quantitative analysis with qualitative information base on interviews with PCCB regional bureau chiefs. Questionnaires were administered in 7 regional offices by EcomResearch – a Dar es Salaam based research firm. Questions related to participation in training and the perceived effectiveness of that training. Questions also dealt with remaining obstacles to PCCB investigations and prosecutions.

While our evaluation focuses primarily on the training activities provided by ICITAP; it is possible that ICITAP's influence extended beyond these training sessions. ICITAP officials particularly point to the technical advice offered by their intermittent technical advisor to the PCCB as a source of influence. Interviews with program participants indicate a strong relationship between ICITAP and the PCCB. However, we are unable to document the extent of ICITAP's influence over the PCCB through their technical advisor. We instead focus on the training programs, as these offer the greatest hope of quantitatively assessing ICITAP's role.

Outputs

In the course of its activities, ICITAP provided all outputs described in the background section above.

Outcomes

QUANTITATIVE RESULTS

In this section, we investigate the effect of training on PCCB performance. We first examine the effect of training on the ratio of convictions to total court cases. Training may boost this ratio by ensuring that those prosecutors have access to more complete and compelling evidence of defendants' culpability. It may also ensure that fewer instances are brought to trial where no such culpability exists. Recall, from the claims above, an effect of training

⁴²Most training sessions included at least a few members of the PCCB headquarters staff. The training in criminal records management software only targeted members of the headquarters staff.

may be present even if no such change in the ratio of convictions to court cases can be found. If there is no change in this ratio in response to training, we would expect to see the fraction of investigations leading to trial or the fraction of referrals leading to an investigation rise.

Figure 2 presents a graphical representation of the coefficients of interest and their 95 percent confidence intervals. Coefficient estimates represent the difference in the average rate of change in the ratio of convictions to court cases in regions in which all districts received training to those in which no districts received training. As a result of discussions with the PCCB regarding the amount of time necessary to investigate a case and bring it to trial, we test for such differences 6 months to one year after training is provided. If training increases the ratio of convictions to on-going court cases, the coefficient estimates presented below would be greater than zero and the 95 percent confidence intervals would not cross the zero line.

As is clear from Figure 2, none of the coefficient estimates are close to significance at the 95 percent level. Nor are the directions of the coefficient estimates consistent – they are negative at 6 months and after 10 months and positive from 7 to 9 months. In all cases, the magnitude of the coefficient estimates is close to zero. We cannot conclude, based on this evidence, that ICITAP training affected the ratio of convictions secured to prosecutions advanced.

It remains possible, however, that training affected the PCCB's performance – and thus prosecutorial behavior – in other ways. For instance, the greater technical ability of the PCCB may imply that prosecutors are now able to bring a wider variety of cases to trial than previously. So while the ratio of convictions to court cases remains roughly constant, the ratio of cases brought to court may increase.

In Tanzania, many more serious corruption cases are prosecuted by the DPP. The PCCB has jurisdiction to prosecute corruption only in cases of bribery. All other corruption prosecutions fall under the jurisdiction of the DPP – though the DPP may delegate authority to prosecute these cases to the PCCB. We therefore examine the effect of training on two ratios – the ratio of cases brought to court relative to the total number of open cases, and the ratio of cases referred to the DPP to the total number of open cases. An increase in either or both of these ratios would indicate that prosecutors are more likely to bring a case to trial if that case was worked in regions where many district offices received training. This would suggest that training increased the technical capabilities of the PCCB, such that it is able to present sufficient evidence to bring more difficult corruption cases to trial.

Results from these regressions are reported in Figures 3 and 4. The specifications used to obtain these coefficient estimates are identical to those discussed above. Our specifications allow for training to affect the ratio of prosecutions to total cases between four and eight months after training is first administered.

As is evident from Figures 3 and 4, in no instance is the coefficient on training significant. That is to say, we are never 95 percent sure that the rate of change in the ratio of cases brought to prosecution (either by the DPP or the PCCB) to total cases is significantly different in regions where more districts received training and regions where fewer districts received training. The sign of the coefficient estimates is also inconsistent. We cannot conclude, based on this evidence, that ICITAP training affected the ratio of prosecutions to cases investigated.

One final possibility is that ICITAP training altered the incentives of the PCCB to undertake an investigation in the first place. Perhaps the ratio of prosecutions to cases investigated is unchanged because the PCCB began to undertake more difficult investigations after training – investigations that are less likely to lead to prosecutions. Both the distribution of cases investigated and the distribution of cases brought to court changed as a result of training – though the ratios of prosecutions to open cases and convictions to court cases remained unchanged.

If this explanation is correct, we would expect to see a change in the ratio of new cases investigated by the PCCB to the number of referrals received. We test for this relationship using an identical specification to those used above. The amount of time necessary for training to affect this ratio is considerably less than that in the above tests – since there is no need to complete an investigation or prosecution before the effects of training are realized. We therefore test for an effect between one and six months after training is administered.

The results of this regression is presented graphically in Figure 5.

As is evident from Figure 5, little consistent pattern of an increase in the ratio of new cases to referrals is evident. There is a marginally significant *decline* in the ratio of new cases to referrals one month after training. But this result is most likely the product of chance. Even if the true value of all coefficients is zero, the probability of observing at least one significant coefficient when six coefficients are estimated is 0.26. Moreover, the direction of the coefficients is inconsistent across months.

It is possible that while training did not have an effect on PCCB performance in general, certain types of training were effective. Training in internal affairs and in supervision

and management seem unlikely to directly affect PCCB investigations. Such training may improve PCCB performance in other dimensions, but is not directly related to the quality of investigations.⁴³ Training in economic and financial crimes or fraud, on the other hand, directly pertains to the quality of PCCB investigations. We therefore test the effect of these rounds of training, decomposing the effects of advanced, intermediate and basic training.

Details of our analysis of the effects of specific rounds of training are reported in section A.2 of the Appendix. These results do not indicate that any of the training sessions examined had a significant effect of PCCB performance.

QUALITATIVE RESULTS

Of six PCCB regional bureau chiefs interviewed,^{44,45} the majority report that they participated in the training provided by ICITAP.⁴⁶ Several did not recall whether the training was effective or not. PCCB training sessions are sufficiently frequent that it is difficult to differentiate the effectiveness of one session from another. Though those that did recall these sessions report that they were useful.⁴⁷

Approximately half of participants report that explicit attempts were made to convey the lessons of training from participants to non-participants. The absence of an explicit system may have diminished the effectiveness of ICITAP training in districts where such a system was absent. Moreover, the four bureau chiefs with the clearest recollection of the ICITAP training program report concerns that training sessions were too short, that an insufficient number of staff were trained, and that seminars were insufficiently catered to the Tanzanian legal system. One bureau chief also suggested that greater attempts be made to incorporate more practical lessons with modern equipment.

All report improvements in investigative capabilities over time – though we lack infor-

⁴³It is, of course, possible that such training may have an indirect effect on the quality of investigations. Better internal affairs supervision may deter corruption amongst investigators – leading to more successful investigations of cases. Or it may lead to the replacement of poor staff. However, such processes would likely require more time to manifest themselves than would direct training in investigative methods. It is also possible that such training would have positive spill-overs across districts which would not be identified using our difference-in-differences methodology.

⁴⁴Seven interviews were intended. One PCCB Bureau Chief was not available due to the requirements of duties pertaining to the upcoming national elections. Interviews took place in September and October of 2010 – two years after the close of the TCP.

⁴⁵The PCCB is subdivided into regional and then into district-level offices. We interview regional bureau chiefs – the head administrators for each regional PCCB force. Tanzania is divided into 26 regions, 23 of which are on the Mainland.

⁴⁶Several also report participating in NYC DoI sessions.

⁴⁷2 of the 6 Bureau Chiefs did not recall the content of training or details regarding its usefulness.

mation to corroborate these claims. Such improvements may reflect many factors; though it is possible that ICITAP and other TCP participants contributed to this improvement.

Moreover, nearly all respondents indicate that the judicial system is now the greatest obstacle to prosecuting complex cases based on circumstantial evidence. This obstacle may affect our quantitative results. If local magistrates are unwilling to convict offenders based on circumstantial evidence, training may not improve conviction rates. No matter the improvement in investigative skills, judges may choose to disregard circumstantial evidence. Moreover, if the DPP and PCCB are less willing to prosecute such cases in light of the judiciary's reticence to accept circumstantial evidence, then the number of cases brought for prosecution may not increase with training regardless of any improvement in investigative skills.

Sustainability

There are reasons to be concerned with the sustainability of any gains reaped from the ICITAP training program. Some efforts were taken to increase sustainability – namely, ICITAP trained PCCB investigators to act as future instructors for training seminars.

However, the short duration of these training sessions and lack of follow-on training raises questions over the amount of information that will be retained by participants in the long term. Most of these training sessions took place over a relatively brief period, 10-11 days. Given the concerns expressed by the bureau chiefs with the duration of training, the prospects for sustainability may be affected. Many regional bureau chiefs do not report any institutionalized attempts to reinforce the lessons of training after the fact, which suggests that any benefits from training may diminish over time.

However, the training related to criminal records management software (CRMS) and to internal affairs appears to have proved more sustainable. ICITAP provided training to PCCB officials to help them develop and refine their CRMS system. This training focused on the use of open source software with which records of criminal investigations could be maintained and organized. The PCCB continues to use such a system today.

There are anecdotal reports that the internal affairs training provided to the PCCB was sustainable. This training involved more extensive preparation than did many of the criminal investigation training sessions. Two technical advisors provided three weeks of assistance before training began. The Director General of the PCCB, noted that the internal affairs training was among the most effective activities undertaken by ICITAP. Moreover, he reported that the PCCB's internal affairs regulations were significantly strengthened

following ICITAP training.

Summary of Results

Based on the above results, our conclusions regarding the effectiveness of the ICITAP training are as follows:

- We cannot conclude that any form of training provided by ICITAP increased the percentage of on-going court cases leading to convictions.
- We cannot conclude that any form of training provided by ICITAP increased the willingness of prosecutors to bring cases to court or the willingness of the PCCB to launch a case.

NYC DoI Training

Background

In a late addition to the Tanzania TCP, USAID approached the New York City Department of Investigations (NYC DoI) to provide additional training in investigative methods to members of the PCCB.⁴⁸ A \$75,000 budget was allocated for logistics related to this intervention. USAID sought NYC DoI's assistance at the request of the PCCB, which had a long history of contact with NYC DoI dating to 1999. NYC DoI previously provided training and equipment to the PCCB – including exchange training sessions in which PCCB officials would directly observe the workings of NYC DoI in New York. In this instance, Mr. Hoseah recommended contacting NYC DoI to provide training in investigative methods – particularly methods related to corruption in the construction industry.

The training provided by NYC DoI consisted of two elements:

- NYC DoI conducted a two week-long training seminar on surveillance methods, report writing, and general elements of corruption investigations with PCCB staff.
- NYC DoI staff were seconded to the PCCB to take part in two ongoing investigations side-by-side with PCCB staff.⁴⁹

⁴⁸The contract to provide logistical services for this training was awarded to Kilimanjaro International in April of 2007.

⁴⁹The most prominent of the two investigations involving NYC DoI focused on the building of the Bank of Tanzania twin towers. The construction of these two towers – located in central Dar es Salaam – involved expenses above and beyond what was required for the offices of a central bank. Total costs ran to \$340

Methodology

It is difficult to quantify the successes or failures of NYC DoI's training efforts. The seminar sessions included 60 officials selected non-randomly from across Tanzania. A session was offered only once and in a single two-week period (rather than in multiple instances for subsets of trainees). Therefore, this report undertakes a qualitative evaluation of NYC DoI's activities.

The evaluation then considers the relation between intended and achieved outputs and attempts to deduce the likely impact from NYC DoI's activities, based on information provided by program participants.

Outputs

The NYC DoI delivered all outputs described in the background section above.

Outcomes and Sustainability

The long-standing nature of the relationship between NYC DoI and the PCCB increased the likelihood that training efforts were successful. In interviews, both the director of the PCCB and the deputy commissioner of the NYC DoI pointed to previous contact as laying a sound basis for this intervention. Past experiences likely allowed NYC DoI to develop a firm grasp of the PCCB's strengths and weaknesses. As such, it was well placed to design a program that was catered to the PCCB's needs.

USAID was also particularly impressed by the cooperation between NYC DoI and the PCCB during joint investigations.⁵⁰ These investigations had a direct effect on corruption.⁵¹ Also – and more importantly for sustainability – program participants indicated that these investigations had an indirect effect through the transfer of skills to PCCB employees. The joint activities involved skills clearly pertinent to the PCCB's investigatory duties. And these skills were employed in real cases rather than in lectures or attempted simulations. Moreover, NYC DoI's advice was likely to have quick and clear effects, in a manner that cannot be replicated in a seminar room.

Officials at both NYC DoI and USAID anecdotally expressed satisfaction with the progress million, while initial forecasts put the price tag at \$80 million. Suspicious expenditures included the building of heliports on the roof of each tower and the construction of expansive parking facilities, which were not built to code. Several prosecutions were launched as a result of these investigations and are currently ongoing. The then Governor of the Bank of Tanzania resigned – both as a result of this scandal and of another scandal involving fictitious payments of external arrears – and subsequently fled to the United States.

⁵⁰Interview with Program Manager, USAID. November 2009.

⁵¹For example, the successful investigation of the Bank of Tanzania towers scandal.

of the PCCB resulting from this effort. The principal in charge of the NYC DoI program took part in both the seminar training programs and in the Bank of Tanzania towers investigation. He noted that officials that took part in the training seminars were able to file better case reports than those that did not, during the course of that investigation. He also noted that PCCB investigators became better able to focus on specific tasks necessary to develop a case over the course of the PCCB-NYC DoI exchange.⁵²

As noted above, the NYC DoI and the PCCB were involved in a longstanding relationship before the implementation of the Tanzania TCP. All indications are that this relationship continues after the TCP's close. The NYC DoI reports that training activities and personnel exchanges continue.

Summary of Results

- The design of NYC DoI's training benefitted from a relationship between NYC DoI and the PCCB dating back nearly 10 years at the time of the TCP.
- NYC DoI staff participated actively alongside PCCB investigators in two ongoing corruption investigations. One of these investigations⁵³ led to very high profile charges and prosecutions. The Bank of Tanzania Towers case alone involved some \$340 million in questioned expenditures.
- The NYC DoI and PCCB joint investigations anecdotally led to skill transfers between the two agencies.
- The relationship between the NYC DoI and the PCCB remains active, reinforcing any gains made during the training sessions conducted as part of the TCP.

Crown Agents Training

Background

Crown Agents was awarded \$146,000 to provide specialized training in procurement practices to PCCB officials. It was understood that the PCCB lacked sufficient familiarity with procurement procedures, and that such training would enhance its ability to prosecute corruption in procurement – in keeping with the stated aims of the agreement.

⁵²Interview with the Deputy Director of the New York City Department of Investigations. June 2010.

⁵³The Bank of Tanzania Towers investigation.

Two training sessions of two weeks each were to be provided for 50 PCCB staff. Topics were to include: Corruption Prevention Strategies; Procurement Practices; Risk Management and Control in Procurement; Audits; Investigating Malpractice; and Ethics and Corruption in Procurement.

Methodology

We conduct a qualitative review of the training provided by Crown Agents Consulting to the PCCB. We attempt to assess the logical link between this training and the reduction of corruption. We also rely on document review and interviews with program participants.

Outputs

Crown Agents Consulting met all outputs.

Outcomes and Sustainability

The purpose of this intervention was well-targeted. Individuals – for instance the PCCB bureau chiefs interviewed above – widely report procurement as a potential area of concern with regards to corruption. It is reasonable to suspect that a working knowledge of the procurement process is vital for PCCB officials to adequately detect corruption.

Moreover, Crown Agents coordinated its training efforts with officials at the Public Procurement Regulatory Authority (PPRA). PPRA officials helped Crown Agents select the trainers for this intervention and also provided guidelines for the materials that would be covered.

However, the training provided by Crown Agents, like that provided by ICITAP, was likely too short and too removed from the day-to-day activities of PCCB agents to have a significant effect on their performance, sustained or otherwise. As noted above, we were unable to isolate a quantitative effect from ICITAP training. And PCCB bureau chiefs pointed to the length of training as a significant problem. Both the duration and the methodology of the Crown Agents training mirrored that provided by ICITAP. There is therefore reason to suspect that similar results were produced. However, we do not have sufficient evidence to corroborate this possibility. Without quantitative evidence of the effectiveness of training it is impossible to say conclusively whether or not this training was effective.

The training provided by Crown Agents is not currently being continued and training did not incorporate a train-the-trainer methodology. It therefore seems improbable that

any gains from Crown Agents' training have been sustained.

Summary of Results

- Training was fully implemented and the number of trainees exceeded expectations (by 10 participants).
- The training was well-targeted and Crown Agents engaged the PPRA in designing training.
- However, training was of a short duration. There were also problems of attendance.

3.2.2 Training of the Department of Public Prosecution

Background

The Department of Public Prosecution (DPP) is the authority charged with most corruption-related prosecutions in Tanzania. Prior to 2007, the DPP was confined to trying a small number of major cases. All other prosecutions would be handled by other authorities, such as the PCCB or the police force. In 2007, the National Prosecution Services Act granted the DPP the authority to prosecute most criminal cases (including corruption cases) from the initial stages of the prosecution. Since initial prosecutions were earlier primarily tried by the police forces, this process was referred to as the 'civilianization' of the prosecutorial service.

Civilianization necessitated the hiring of a large number of new DPP staff. Many of these recruits were new law school graduates. And the DPP discovered that knowledge of the Prevention and Combating of Corruption Act of 2007 amongst these new recruits was lacking.

USAID amended the Tanzania TCP to include training for these new recruits to the DPP, in light of the reforms then taking place. A contract was awarded to Kilimanjaro International for \$609,000 in December 2007. Eight training sessions, of three days or one week each, were to be provided to 157 new recruits to the DPP. The training sessions included: Criminal Prosecution Skills; Writing Legal Opinions; Information and Communications Technology; Trafficking and Terrorism; Human Rights, Gender, and Vulnerable and Disadvantaged Groups; Cyber-crime, Money Laundering and Fraud; Management; and Fighting Corruption.

Methodology

Because all recruits were trained simultaneously and because this training was contemporaneous with institutional changes at the DPP, it is not feasible to quantitatively identify the causal effect of this activity. Any comparisons of the DPP's activities before and after this program would reflect the effects of the whole 'civilianization' process, and not those of training. Moreover, because training was provided to new recruits, we cannot compare the performance of the same individuals before and after training was provided.

Rather than attempting to quantify the effect of training, we qualitatively assess the relation between training and the reduction of corruption. To do this, we examine the implementation of the program – relying on information provided by TCP participants.

Outputs

All outputs were delivered as intended.

Outcomes and Sustainability

Program participants indicate that this program was well-implemented and successful in improving corruption prosecutions, relative to the counterfactual case in which no training was provided. There are several aspects of this program that led to these reports of success:

First, the training provided in this instance was a response to a precise need documented by the DPP. The DPP itself documented the weaknesses of its recruits. This included a lack of a working knowledge of the aspects of the new Prevention and Combating of Corruption Act, as well as a lack of practical experience in the tasks that regularly face prosecutors. The DPP is naturally knowledgeable of the demands facing new recruits, and is thus it was well-placed to determine the types of training they would require.

The training program was designed to cover the topics noted as necessary by the DPP. To address the Prevention and Combating of Corruption Act, the training program included a three day seminar largely focused on this Act. Since the Act expanded the set of charges that might be brought for corruption, much of the focus was on the content of these new charges. The program also included training in practical matters – such as drafting legal opinions and working with information technology – that the DPP noticed were inadequately covered by law school curricula.⁵⁴

⁵⁴These sessions, however, only covered a brief period of time. We do not possess adequate information to determine whether or not the duration of the training was sufficient to adequately cover the topics deemed necessary.

Second, the intervention was a response to institutional reforms brought about by the Government of Tanzania. Since training was provided to what was, in essence, a largely new institution, it was particularly influential. Trainees were not seasoned veterans with long-established ways of conducting business. Rather they were new recruits to an institution assuming a new role. Patterns of behavior were thus likely to be malleable, and amenable to the input of trainers.

Third, the Government of Tanzania continues to fund training for new recruits to the DPP; although the training regimen has been modified. The willingness of the Government to assume this role indicates that the TCP training was perceived as valuable. This suggests that training was effective. This also bodes well for the sustainability of the training. However, the DPP is looking for external support for additional training activities.

Summary of Results

- The training of 157 new recruits (the full set of all recruits) to the DPP in 2007.
- The training new recruits contributed to a major institutional change at the DPP.
- Because training was provided to a large class of new recruits when the DPP was assuming a new role, and because training was tailored to meet specific requests of the DPP, program participants perceived this intervention as effective.

3.2.3 Legal Aid

Background

The legal aid intervention was intended to improve the access of Tanzanian citizens to the judicial arena. Citizens lacking the resources otherwise necessary to seek legal redress for others' malfeasance could instead take their case to a legal aid clinic to receive legal advice and help preparing legal documents.⁵⁵ To the extent that citizen complaints relate to corruption, improvements in the provision of legal aid would facilitate the public's ability to monitor and punish said corruption through legal mechanisms. Legal aid clinics might help direct citizens to the appropriate authorities to file a complaint, or they may directly aid citizens in bringing a corruption-related case to court.

⁵⁵It is worth noting that the legal aid clinics generally keep at least one lawyer on staff. However, they do not keep an advocate (equivalent to a barrister in the British legal system). Only an advocate can represent a client in court.

To meet these goals, USAID contracted Women in Law and Development in Africa (WiLDaf) under a \$640,000 grant. WiLDaf undertook a number of interventions broadly aimed at improving the capabilities of legal aid providers and at coordinating the activities of these providers.

Methodology

Unfortunately, we lack data on the number of individuals seeking legal aid provision and the outcomes of their cases. Without such data, both before and after WiLDaf's training efforts, we cannot quantify the effect of WiLDaf's training on legal aid provision. Nor can we assess whether or not the media program devised by WiLDaf was successful in increasing the number of eligible individuals seeking relief through legal aid.

We therefore rely on a qualitative evaluation of WiLDaf's efforts. We assess the relationship between these efforts and the overall goals of the Tanzania TCP.

Outputs

Over the course of the Tanzania TCP, WiLDaf accomplished the following:

- Creation of a legal aid secretariat involving 8 of the 12 major legal aid providers in Tanzania.
- Creation of five model legal aid clinics.⁵⁶
- Publication of two pamphlets on legal aid and corruption and convening several workshops for 150 legal aid providers.
- Launching of a radio campaign to raise awareness of legal aid provision.
- Launch of a 16 days of Activism Against Gender Violence campaign.

Outcomes

While the adequate provision of legal aid has numerous possible benefits; the logical link between this intervention and the effective monitoring of corruption is tenuous. There is reason to expect that an independent judiciary may act as a check on corrupt activities. But it is less clear that improving access to legal aid will improve the functioning of a judicial check on corruption.

⁵⁶Though the number of clients at these clinics fell slightly short of expectations.

First, it is not clear that those eligible for legal aid are likely targets for corrupt officials. Legal aid is intended to provide support for the poor, particularly in rural communities. However, corrupt officials are more likely to target members of the business community as sources of bribes than they are to target poor and rural communities – as members of this community have a greater demand for state services and greater ability to pay.⁵⁷ Any demands for bribes from the rural poor are likely to be of insufficient amounts to render legal action a viable response.

Second, it is unclear that the type of corruption most likely to affect those eligible for legal aid is most amenable to individual legal action. It is unlikely that incidences of petty bribery will involve sufficient sums to be worth the cost of litigation. If corrupt politicians are misallocating funds from local budgets, individual citizens are unlikely to possess the investigative resources necessary to pursue a legal remedy by themselves.⁵⁸

Third, it is unclear whether the costs and difficulties of litigation are the primary barriers preventing access to the legal mechanism. Indeed, WiLDaf reports that many individuals with claims of corruption do not come forward because they fear retaliation.⁵⁹ The provision of legal aid is unlikely to alleviate that fear.

Moreover, even if these challenges to the logic of the intervention are incorrect, the majority of the activities carried out by WiLDaf did not specifically focus on corruption. While training was provided to legal aid staff on the content of the Prevention and Combating of Corruption Act of 2007 – and was coordinated with the PCCB; the majority of activities aimed either at increasing public awareness of legal aid services or improving the capacity of legal aid services. This focus is certainly understandable given WiLDaf's mandate. But it suggests that this intervention did not bear a clear logical relation to the rest of the TCP, which focused on combating corruption.

The majority of these activities bear a clear logical relation to the provision of legal aid in Tanzania. And all accounts indicate that the intended program outputs were delivered – although we lack adequate records to assess performance beyond outputs.

⁵⁷?, for instance, find evidence of price discrimination in the solicitation of bribes from truckers in Aceh. Corrupt officials seek more substantial bribes from those whose observable characteristics indicate a higher ability to pay. However, it is worth noting that the poor may be victimized through the reduced provision of public services resulting from corruption, as corrupt officials steer services towards those best able to pay.

⁵⁸Moreover, legal aid clinics would be unable to represent citizens in such suits. As with many common law systems, Tanzania features a split legal profession. Advocates may represent clients in court, lawyers may provide advice outside of court. Legal aid offices include lawyers, but not advocates. See the Macro International Final Performance Audit, p. 24.

⁵⁹Final Performance Audit MCA/Tanzania Threshold Country Program. Macro International. June 2008.

Sustainability

The institution-building activities pursued by WiLDAf produced bodies that function to this day. Both the model legal aid clinics and the legal aid secretariat continued to operate after the close of the Tanzania TCP.

However, training activities ceased for a time. Training resumed in 2009 – after the close of the TCP – with the support of USAID.⁶⁰ This may be a concern for the sustainability of this intervention.

There are causes for concern regarding the sustainability of awareness activities. The success of such campaigns is conditional on the degree of exposure (Barabas and Jerit, 2010). We do not possess information as to whether WiLDAf continued its advertisements. Though the 16 Days Against Gender Violence was a one-off event.

Summary of Results

- WiLDAf successfully completed all outputs.
- However, WiLDAf's activities bear a tenuous relation to the aims of the TCP, namely the combating of corruption.
- WiLDAf's building of institutions – i.e., the legal aid secretariat and the 5 model legal aid clinics – have produced institutions that continue to operate.
- However, we lack adequate data to fully assess the effectiveness of WiLDAf's activities.

3.3 Financial Intelligence Unit

Background

The third component of the TCP called for the establishment of a Financial Intelligence Unit (FIU), for the purpose of cracking down on money laundering and the financial support of terrorism. The FIU – coupled with the passage of an Anti-Money Laundering Act (AMLA) in 2006 – was intended to reduce the ability of corrupt parties to move and disguise the proceeds of corruption. To achieve this, the FIU was intended to:

⁶⁰Interview with Judith Odunga, WiLDAf. April 2010.

- receive suspicious transaction reports (STRs) from financial institutions and to analyze and disclose these reports to the proper investigatory authorities,
- track information on the number of investigations and prosecutions under the AMLA,
- compile statistics on confiscated funds and assets and to provide these figures to the public.

Deloitte & Touche was awarded nearly \$1.2 million to support the creation of an FIU and to secure training for its officials.

Methodology

It is quite difficult to quantify the effect of the FIU's activities. This difficulty arises for two reasons: First, the intended outcome is, by its very nature, difficult to adequately measure. Money laundering is intended to evade detection – it is therefore difficult to devise a valid measure of money laundering. Second, we do not have an adequate control group with which to compare the results of the intervention – the FIU is an institution that affects mainland Tanzania as a whole.⁶¹ Even if we had a measure of money laundering activity, the only valid comparison we could perform would be over time. Such a comparison would *not* capture the causal effect of the creation of the FIU, as results would be confounded by other events that may effect the prevalence of money laundering.

Instead, we assess the logical relationship between the intervention and the outcome of interest. We also examine the records regarding the implementation of this intervention and prospects for achieving the desired results. To do this, we rely on interviews with program participants, materials produced during the course of the TCP, and on a mutual evaluation report on the Tanzanian FIU released by the Eastern and Southern Africa Anti-Money Laundering Group (ESAAMLG) in December of 2009.

Output

- FIU was established. However, it was only partially operational by the close of the TCP.
- Members of the financial sector were informed of their reporting obligations under the new anti-money laundering law.

⁶¹The FIU's mandate does not extend to Zanzibar. Though Zanzibar did pass an anti-money laundering bill at the same time as the Anti-Money Laundering Act of 2006 came into force.

Outcomes

The FIU was intended to serve as an information clearing house regarding money laundering. The creation of a central institution responsible for the collection and provision of information about money laundering bears a clear logical relation to the purpose of lowering levels of such laundering and reducing corruption. It reduces potential barriers that might emerge between financial regulators and police forces. It establishes clear lines of responsibility in the field of policing money laundering. And, the reduction of money laundering may inhibit the conduct of large scale corruption, as guilty parties are less able to hide the proceeds of corrupt activities.

However, several notable concerns – both with program logic and with implementation – lead us to raise questions about the effectiveness of the FIU intervention. First, it is important to emphasize that the ability of the FIU to aggregate information on money laundering is contingent on the willingness of financial institutions to adequately file suspicious transaction reports (STRs). While there is good reason to delegate the authority for unearthing suspicious transactions to the financial sector – these institutions have greater expertise in financial transactions than any government agency, and are more familiar with their consumers' normal financial activities and thus better able to notice a deviation from the norm – this delegation also poses problems. Monitoring financial transactions for possible money laundering imposes a cost on financial institutions, which they may be reluctant to bear.⁶² The disclosure of such information may also harm financial institutions' relationships with their clients and lead to the loss of business. Since the FIU is not authorized to investigate and punish financial institutions for failure to report suspicious transactions (this authority is vested in the Bank of Tanzania), the incentive to properly report suspicious activities may be weakened.

Second, the effectiveness of the FIU may be hampered by forces outside of its control. It relies on police agencies to act on any evidence of money laundering it uncovers. If these agencies are unable to adequately act on these reports – or if prosecutors and/or the judiciary are unwilling or unable to secure convictions for accused money launderers – the effect of the FIU will be diminished.

Third, this intervention suffered from several problems of implementation. Most notably, it was subject to repeated delays. Deloitte began work in January 2008 – i.e., in the final year of the TCP – due to the need to await the Parliamentary confirmation of the

⁶²Indeed, Macro International's Final Performance Audit notes that "Banks have not been eager to disclose information about the suspicious transactions of their clients..." p. 47.

Commissioner of the FIU. Additional delays resulted from difficulties in finding a contractor suitable for the assignment and amenable to the FIU.⁶³ By the time the TCP had come to a close, the FIU had yet to start receiving and analyzing STRs.⁶⁴ However, offices had been created for the FIU, several staff were in place, and two sensitization sessions had been conducted with members of the financial community under TCP auspices.

During evaluation interviews,⁶⁵ the FIU had begun to receive suspicious transaction reports, but still only had 8 staff other than the Commissioner (including secretarial staff and drivers), all of whom had been seconded either from the Bank of Tanzania or from Ministry of Finance and Economic Affairs. The rate of submission of STRs had increased between the November 2009 and April 2010 interviews – from single digits to double digits per month.⁶⁶

Some progress has been made on staffing issues. The FIU's budget included funding for up to 15 new hires. At the time of our interview, the selection process for these positions was ongoing.⁶⁷

Finally, several conflicts emerged in the relationship between the FIU and Deloitte, indicating that the implementer and the agency with which it was partnered had conflicting views of their roles. These problems related to the role of Deloitte in training FIU staff, and are listed below:

- Part of the training program for FIU staff included two attachments to the South African Financial Intelligence Center and to the Financial Crimes Enforcement Network (FinCEN) in the US. Deloitte did not participate in meetings between the FIU and FinCEN and thus was not able to play an active role beyond facilitating this portion of the training of FIU staff. Deloitte indicated its belief that, had it been allowed to participate in the training sessions, it could have played a more constructive role in advising the FIU.⁶⁸

⁶³Interviews with the Commissioner of the FIU and Program Manager, USAID – November 2009 and April 2010.

⁶⁴Final Performance Audit MCA/Threshold Program Tanzania. Macro International. June 2008.

⁶⁵In November 2009 and April 2010.

⁶⁶The ESAAMLG Report listed the previous level of STRs as low, relative to the size of the economy and banking sector. The rate currently being reported is comparable to that in Botswana – though the ESAAMLG also criticizes this rate as low. By contrast, South Africa – with a far larger economy and banking sector – received 24585 STRs in FY 2007/8.

⁶⁷The Commissioner reported that finding qualified recruits was difficult, but was optimistic that new staff would be added to monitor and analyze the STRs, and in the legal and financial fields. Interview with the Commissioner of the FIU. April 2010.

⁶⁸Macro International Performance Audit.

- Deloitte did not develop a relationship with the Bank of Tanzania Training Institute or the University of Dar es Salaam early in the implementation process. Since these two bodies were responsible for the initial training of FIU staff, it is possible that the absence of such a relationship may have hindered subsequent training efforts.

Given the difficulties in staffing, the slow progress in establishing an STR review system, and the reliance of the FIU on cooperation from the financial sector and other agencies, it is unlikely that the FIU had an effect on corruption during the course of the Tanzania TCP or by the time of the evaluation.

Sustainability

In December 2009, the Eastern and Southern Africa Anti-Money Laundering Group (ESAAMLG) published a Mutual Evaluation Report on the activities of the FIU to date. ESAAMLG noted a number of shortcomings in the FIU's activities. In addition to the staffing issues discussed above, ESAAMLG found that the FIU's efforts were hindered by its limited authority over non-bank financial institutions and that the FIU's authority did not extend to Zanzibar.

During our interview in April 2010, the FIU seemed quite responsive to these criticisms. At the time, discussions were underway with capital markets, insurance, and gaming regulators to extend the enforcement provisions of the AMLA to non-banking financial entities. Discussions were also underway with Zanzibar to amend its anti-money laundering law to cross-reference the FIU, expanding the FIU's authority to collect information on suspicious transactions in Zanzibar. Such efforts are difficult to coordinate given the federal nature of Tanzania's government and the sensitivity of Union issues.

Progress in amending the AMLA is unlikely to be great this year, however. While the FIU notes that it receives a great deal of support from the government; it also notes that 2010 is an election year. Substantial legislative changes are unlikely to take place while campaigning is ongoing.

Summary of Results

Based on our qualitative evaluation, we make the following conclusions:

- The creation of the FIU suffered greatly from problems in securing adequate staff and establishing a system for reviewing STRs.

- The delegation of reporting responsibility to the financial sector raises the possibility that non-compliance will undermine efforts by the FIU.
- In practice, the time frame for the creation of the FIU was compressed due to delays in the confirmation of the Chairman and the selection of an implementing partner.
- Given the delays in rendering the FIU operational, it is implausible that its creation had an effect on corruption during the TCP and unlikely that any effect was manifested before the time of this evaluation. It is possible, however, that the FIU will have an effect on corruption levels over the longer term.

3.4 Nongovernmental Sector Monitoring Capacity

The fourth component of the Tanzania TCP sought to build the monitoring capability of the nongovernmental sector. It was thought that increasing the ability of nongovernmental entities to monitor and publicize corruption in the public sector would improve accountability and reduce corruption. The public would be better able to punish elected politicians for corruption in the public sector, increasing pressure on the government to police and punish corrupt activities. And the public would be more aware of corrupt activities and would be better able to observe and report incidents of corruption to the proper authorities.

3.4.1 Investigative Journalist Training

Background

One of the two core aims of the Nongovernmental Sector component of the TCP was to “Train journalists in investigative reporting.”⁶⁹ The purpose of this training was to improve the quality and quantity of reporting on corruption in Tanzania, in the hope that better access to information might improve the ability of citizens to hold officials accountable for corrupt activities.

PACT-Tanzania (PACT) was awarded a \$185,000 grant to oversee the implementation of the investigative journalist training program. PACT was to run a series of journalist training sessions between January 2007 and November 2008, several of which were conducted jointly with the Media Institute of Southern Africa (MISA). Training sessions would

⁶⁹Strategic Objective Grant Agreement between the United States of America and the United Republic of Tanzania to Combat Corruption. May 3, 2006.

focus on specific aspects of corruption and good governance (e.g., public procurement, natural resources management) and were meant to help journalists develop an expertise in investigating instances of corruption or poor governance in these areas.

During the period between January 2007 and November 2008, 10 different types of training sessions were provided: training in corruption and good governance, in public expenditure tracking (PETS), in procurement regulations, in natural resources management, training provided by the Media Institute of Southern Africa (MISA), training provided by veteran foreign journalist Jonathan Powers, and regional training sessions in the Lake Zone, Southern Zone, Southern Highlands, and Northern Zone. Most training sessions consisted of 3-4 days of seminars on the particular topic of focus. However, the Powers training session also involved the provision of travel grants to participants to develop a story currently in the works. After returning from their travels, participants and trainers would jointly criticize the stories under development, with the aim of improving the quality of the work produced.⁷⁰

In total, 312 participants took part in training, 22 of whom attended more than one session. In addition to journalists, several editors took part in training. And some journalists who took part in training subsequently became editors.

Methodology

For the training of journalists in investigative methods to have an effect on public perceptions of corruption and – ultimately – on the accountability of public officials, this training must produce at least one of two results. It may increase the number of investigative articles published, thereby increasing public exposure to corruption-related issues and uncovering a larger number of corruption cases. Or, it may increase the quality of coverage – improving public understanding of the mechanisms by which corruption operates and/or providing more credible evidence of corruption. In our evaluation, we test for both possible effects.

THE EFFECT OF TRAINING ON THE QUANTITY OF ARTICLES PUBLISHED

To assess the effect of training on the number of articles produced, we make use of a collection of articles PACT compiled from Dar es Salaam circulated newspapers from January 2006-November 2008. Two months during this period – April and May 2008 – are missing

⁷⁰Several participants from earlier sessions were also invited to participate in the travel grant program and the subsequent critiquing. Only journalists in the Powers session, however, had 100 percent participation in these elements of training.

from the dataset, as the binders containing the original press clippings were misplaced in PACT's offices. Articles are collected from 36 newspapers, the titles of which are enumerated in Table A.2 in the Appendix. In total, there are 3,297 articles listed by 991 separate authors.^{71,72}

To assess the effect of training on the number of articles produced, we structure the data so that a single author may be observed over time. More precisely, we examine the number of articles published by a given journalist in a given month. The total number of articles produced by each author over the full January 2007-November 2007 period is low, such that most authors in most months do not produce a publication. The average number of publications per month is 0.15. A plot of the distribution of the number of articles published in a given month is presented in Figure A.6 in the Appendix.

A naive estimation strategy might examine the average number of articles produced by those journalists who were trained and compare these to the average number of articles produced by those who were not. However, it is likely the case that authors who received training differed systematically from those that did not. In the early stages of the program, PACT selected journalists for training based on recommendations from editors and from professional organizations. Later, many journalists applied directly for places in the PACT training programs. It is likely that those journalists who were selected for training were more regularly involved in corruption investigations than those that were not. Thus, any difference in the average number of articles produced by trained and non-trained journalists may be attributable to factors other than training.

However, these factors are unlikely to affect the *rate of change* in the number of articles produced per month. Training, if effective, would be expected to increase the rate of change in the number of articles produced by a journalist from the month prior to training to the month after. We therefore focus our analysis on the *change* in the number of articles produced by a given journalist from one month to the next.

Figure A.7, in the Appendix, displays the distribution of the change in the number of articles produced from one month to another. Since, in most months, most journalists do not write a corruption-related article, the most common change is zero. Over the period, the average change in the number of articles produced is very close to zero and slightly

⁷¹This is a list of all articles that include a byline. Between $\frac{1}{3}$ and $\frac{1}{2}$ articles are published anonymously. We discuss the consequences of these anonymous articles for our estimates of the effect of training below.

⁷²These data were reviewed by Macro International during its performance audit of the TCP. Macro reviewed PACT's clippings and compared these to a set of clippings composed by Steadman Group.⁷³ These reviews indicated the validity of the PACT data. Interviews with former Macro staff, June 2010.

negative (-0.005). However, there is considerable variation in the rate of change, from a minimum of -8 to a maximum of 13.

We rely on PACT's records to code which journalists took part in training and when this training took place. Authors are matched by name. In matching these records, we allow for minor differences in the spelling of given and surnames, such that spellings may differ by one letter (e.g., substituting a 'i' for a 'y'). In addition to examining the average effect of training across all seminars, we allow this effect to vary across each type of seminar.⁷⁴

The nature of the data used for this analysis ensures that we offer a conservative estimate of the effect of training. As noted above, many of the articles collected by PACT are published anonymously and these articles are excluded from the data analyzed below. If authors publish some articles anonymously and publish others with bylines, this will tend to bias our results towards zero. The reason for this is simple: Assume that training increases the number of articles published by a given journalist, but that each journalist only includes a byline with a probability of one-half. Then, on average, the observed change in the number of articles published by a trained journalist will only be half as large as the true change. The observed difference between trained and untrained journalists will thus be half as large as the true difference.

Similarly, our data are only drawn from newspapers that are circulated in Dar es Salaam. Since a number of journalists trained by PACT hailed from other regions, it is likely that some of their publications did not reach Dar es Salaam-based papers. If only a fraction of each of these journalists' publications reached Dar es Salaam papers, our results will be biased towards zero for the same reasons discussed above. This bias is particularly likely to affect the training sessions conducted for the Southern, Southern Highlands, and Lake zones.

THE EFFECT OF TRAINING ON THE QUALITY OF ARTICLES PUBLISHED

To assess the effect of training on article quality, we examine a panel of articles produced by randomly selected authors. Authors were selected for this panel via stratified random sampling from the pool of all authors in the dataset described above who had written at least two articles during the January 2007-November 2008 period.⁷⁵ A total of 55 authors were selected for the quality panel, 25 of whom received training from PACT and 30 of

⁷⁴ Unfortunately, however, the Northern Zone batch training took place in May 2008 – a period for which the number of articles published was not recorded. Thus, we cannot assess the effect of this round of training.

⁷⁵ We confine our attention to authors who produced at least two articles since we wish to assess the change in article quality over time.

whom did not. We deliberately over-sampled authors who produced a large number of articles (28 authors in the panel produced seven or more articles over the period), so that we might better estimate changes in article quality over time.

A group of four enumerators from EcomResearch read each article written by authors in the sample. A total of 421 articles were listed in the database compiled based on PACT's article clippings, of which 374 could be retrieved.⁷⁶ Each enumerator was asked to independently respond to a number of subjective and objective questions regarding each piece. The content of these questions is enumerated in Table 5.

Table 5: Quality Questionnaire

Question	Values
Would you rate the overall quality of this article as among the bottom 20% of newspaper articles you have read? (Value 1) As among the 20-40% of newspaper articles you have read? (Value 2) As among the 40-60% of newspaper articles you have read? (Value 3) As among the 60-80% of newspaper articles you have read? (Value 4) As among the 80-100% of newspaper articles you have read? (Value 5)	{1, 2, 3, 4, 5}
Do you think this article would be easily understood by a member of the general public? Do you think he/she would need to reread the article more than once to comprehend the main argument?	{0, 1}
How many grammatical errors or typos, on average, did you find per page?	{0, 1, 2, ..., N}
Does the article cite any sources (by name or anonymously)?	{0, 1}
How many sources does the article cite (by name or anonymously)?	{0, 1, 2, ..., N}
Does the article attempt to relate specific instances of corruption to broader economic or political issues, or does it only concentrate on a specific instance of corruption?	{0, 1}
Is the article written in Swahili?	{0, 1}
Do you think this is a highly technical article (containing multiple technicalities or technical explanations)?	{0, 1}

⁷⁶EcomResearch reported difficulties in locating and reading the remaining articles. An examination of the missing articles suggests that these were missing at random from the dataset. As such, the precision of our estimates is reduced by the missing observations, but the missingness should not induce bias in our estimates.

We developed a set of questions listed in Table 5 that relate to aspects of the quality of reporting. For instance, the use of appropriately cited sources reflects the thoroughness of the investigation carried out by reporters. The ability to place a particular instance of corruption into broader context reflects the informativeness of the article. The ‘overall quality’ indicator helps to pick up on subjective perceptions of article quality that may not be adequately captured in the other variables. Other measures were intended to control for possible biases in the estimates. For instance, the clarity and grammatical quality of the article may vary depending on whether it is printed in English or Swahili. Highly technical articles may prove more difficult to read and may involve more direct citations of experts.⁷⁷ The questionnaire was constructed after consulting relevant literature in political science on media bias, consulting a political communication expert – with a focus on developing countries – at the Annenberg School for Communications, reviewing journalists’ discussion of the quality of American investigative reporting, and discussions with the MCC.⁷⁸

For our estimation of the effect of training on article quality, we construct an index of quality based on the above questions. Our index is constructed using a measurement model based on Item Response Theory (IRT). This method of measurement allows the influence of each question to vary and accounts for differences in the responsiveness of each enumerator. Details on this measurement method, and a figure summarizing the quality measures, are available in section A.3 in the appendix.

We then use this quality index to estimate the effect of training on the quality of publications. We divide the January 2007 to November 2008 period into eight periods (7 last 3 months, one lasts 2 months) and construct a training dummy variable equal to 1 during and after the period in which a given journalist received training. We then assess the following model:

$$QualityIndex_{i,j,t} = \alpha_i + \beta Training_{j,t} + \sum_{k=1}^T \gamma_k \tau_k + \epsilon_{i,j,t}$$

where i denotes journalist i , j denotes article j , and t denotes time t . α_i is a journalist

⁷⁷To test inter-coder reliability, we examine the correlation of enumerator responses to the question “Is the article written in Swahili?” This question involves the least possible room for subjective judgment, and thus should be perfectly or near-perfectly correlated between enumerators. Low levels of correlation would reveal substantial coding errors that would invalidate the data used for this assessment. In fact, the correlation in responses to this question ranges from 0.9715 to 0.9929 across reviewers. Such high correlations suggest that coding errors are minimal.

⁷⁸The MCC provided a tool used to examine media quality by IREX in a separate project.

specific fixed effect and $\sum_{k=1}^T \gamma_k \tau_k$ are a series of period specific fixed effects.⁷⁹

The coefficient β on the training dummy captures the difference in average article quality before and after training for a journalist who receives training during periods two through eight.⁸⁰ The time period fixed effects control for time trends that affect both trained and untrained journalists. This specification, therefore, produces a difference-in-differences estimate of the effect of training.

We also estimate an alternative model adjusting for the partisan nature of the newspapers in which journalists publish. It may be the case that trained journalists that publish their articles in particularly partisan papers tend to use their investigative expertise primarily to attack their partisan opponents, rather than to develop broader lessons regarding corruption and good governance in Tanzania. The effect of training may therefore differ significantly if provided to the partisan, as opposed to the neutral press. We estimate the model:

$$QualityIndex_{i,j,t} = \alpha_i + \beta Training_{j,t} + \delta Pro_Gov_j + \psi Pro_Opp_j + \zeta Pro_Gov_j * Training_{j,t} + \omega Pro_Opp_j * Training_{j,t} + \sum_{k=1}^T \gamma_k \tau_k + \epsilon_{i,j,t}$$

where Pro_Gov_j and Pro_Opp_j are indicator variables that equal one if a given article j appears in either a government-leaning or opposition-leaning newspaper, respectively. We rely on the assessment of the researchers at EcomResearch to define these variables. The partisan affiliations of the newspapers included in this sample are reported in Table A.4 in Appendix A.3.

As a final robustness check, we estimate an alternative model of article quality. We are particularly concerned that a number of journalists in our sample received training in early 2007 and – because of the journalist-level fixed effects – do not influence our estimate of the coefficient on the training term.⁸¹ Yet journalist fixed-effects are necessary to control for the non-random selection of journalists into training. To sidestep this issue, we confine this model to the set of 25 journalists in our sample who received training at some point in time. We then compare the average quality of articles produced by trained journalists

⁷⁹All specifications adjust for sample weighing measures and cluster standard errors by journalist.

⁸⁰For journalists who are trained in period one, the training dummy is perfectly collinear with the journalist-level fixed effect. We do not observe their quality prior to training. Therefore the quality of the articles produced by these journalists does not influence the coefficient on the training identifier.

⁸¹Recall that any journalist trained in the first three months of 2007 will not affect the estimates of this coefficient, since the training variable is perfectly collinear with the journalist fixed effect for these authors.

to that of journalists who had yet to receive training according to the following model:

$$QualityIndex_{i,j,t} = \alpha + \beta Training_{j,t} + \sum_{k=1}^T \gamma_k \tau_k + \epsilon_{i,j,t}$$

where α is common to all journalists. So long as the selection process for journalists who take part in the program was relatively consistent over time, the exclusion of journalist fixed-effects should not prove problematic for this model. The results of this model are reported in Table A.5 in Appendix A.3.

Outputs

PACT delivered all intended outputs as described in the background section above.

Outcomes

THE EFFECT OF TRAINING ON THE QUANTITY OF ARTICLES PUBLISHED

To estimate the causal impact of training, we regress the change in the number of articles published by a given journalist from one month to the next on an indicator variable that takes the value one in the month of training and the value zero in all other months. This type of model is known as a difference-in-differences specification and controls for characteristics of individual journalists that remain fixed over time. The coefficient on the training indicator variable measures the difference in the average change in the number of publications produced by journalists who were trained in a given month relative to those who were not. A positive coefficient indicates that, on average, the change in the number of articles produced by journalists trained in a given month was greater than the change in the number of articles produced by untrained journalists. Such a result would suggest that training had a positive effect on the number of articles produced.

It should be noted that factors other than training may affect the willingness or ability of trained journalists to increase their output. Most notably, the government's reaction to press investigations may tend to diminish the effect of training on journalist output. USAID notes that while the press in Tanzania is free, it is not uncommon for newspapers to have their licenses withdrawn in response to critical press coverage. To the extent that the media was intimidated by such threats, the effect of training on journalist output would be reduced – even were training perfectly effective. However, if training did not affect either the quantity or quality of journalists' output – regardless of whether this was due to

government threats or the ineffectiveness of the training program – it cannot have had an effect on levels of corruption.

Our analysis first focuses on the effect of training averaged across all the different forms of training provided. The variable of interest is, therefore, Δ training, an indicator that takes the value one in the month that a given journalist receives any form of training. We control for a general time trend by including a cubic polynomial of time in the regression.⁸² In some specifications we also allow the effect of training to vary over time by adding measures of the time since training was provided.⁸³ The results are reported in Table 6.

The coefficient on Δ training is positive, but it is not significant at conventional levels. We therefore cannot say that training has a positive contemporaneous effect on the number of articles produced. Nor does it appear that training has a positive lagged effect on the number of articles produced.

While the results above do not reveal a relationship between the number of articles produced and training on average; it remains possible that certain rounds of training were more effective than the others. To test this possibility, we regress the change in the number of articles published by a given journalist against a series of indicators for each form of training provided. These indicators take the value of one if a given journalists received that specific type of training in a given month. A positive coefficient indicates that journalists who received such training experienced a larger increase in the number of articles produced in that month than journalists who did not receive such training. Because of missing data, we cannot test the effects of the Northern Zone batch of training.

These results should be interpreted cautiously, however. In running this regression, we test the significance of nine different indicator variables. Coefficients on these variables are interpreted as significant if the probability of obtaining a coefficient as large or larger than that observed in the data – given that there is no true relationship – is less than one-in-twenty.⁸⁴ The probability that the coefficient on at least one of these indicators is significant, even when there is no relationship between training and the number of articles produced, is $1 - 0.95^9 \approx 0.37$. Because we are making multiple comparisons, the probability that at least one of these comparisons is significant is relatively high.

The results from this regression analysis are presented in Table 7. As is evident from

⁸²Time $t = 1$ in January 2007, $t = 2$ in February 2007, etc. We control for t , t^2 , and t^3 , which allows time effects to assume a very general functional form (Carter and Signorino, 2007).

⁸³In unreported results, we model the lagged effects of training using indicator variables that take the value of one for each month following training, following the methodology of Laporte and Windmeijer (2005). The results are substantively identical to those reported below.

⁸⁴One-in-ten for significance at the 90 percent level.

Table 6: General Effect of Training

	Δ	No. Articles
Δ training	0.079 [-0.051,0.210]	0.080 [-0.050,0.211]
t	-0.007 [-0.017,0.004]	-0.007 [-0.017,0.004]
t ²	0.002*** [0.001,0.003]	0.002** [0.001,0.003]
t ³	-0.000*** [-0.000,-0.000]	-0.000*** [-0.000,-0.000]
Months since training		0.000 [-0.012,0.013]
Months since training ²		0.000 [-0.002,0.003]
Months since training ³		-0.000 [-0.000,0.000]
$\hat{\sigma}^2$	0.717	0.718

Results of a difference-in-differences regression of the change in the number of articles produced by a given journalist in a given month against the change in training status. The estimates pool across all forms of training provided. Controls for a cubic polynomial of time adjust for trend effects. And lagged effects of training are modeled using a cubic polynomial of the number of months since training. 95 percent confidence intervals are reported in brackets below the coefficient estimates. * denotes significance at the 95 percent level, and ** denotes significance at the 95 percent level. $\hat{\sigma}^2$ denotes the standard error of the regression – roughly interpreted as the average difference between the observed outcome and the prediction of the regression. All standard errors are clustered by journalist.

the table, the majority of training sessions are not associated with an increase in the rate of change in the number of articles produced. However, the effect of the training session led by Jonathan Powers is positive and significant at a greater than 99 percent level of confidence. Moreover, the estimated effect of the Powers training is substantively large –

Table 7: Differentiated Effect of Training by Type of Training

	Δ No. Articles
Corruption & Good Governance	-0.079 [-0.287,0.128]
MISA-Tan	0.010 [-0.004,0.024]
PETS	0.095 [-0.228,0.418]
Powers	0.551** [0.198,0.904]
Lake Zone	0.210 [-0.020,0.440]
Natural Resources Management	-0.238 [-0.780,0.304]
Southern Highlands Zone	-0.061 [-0.454,0.333]
Southern Zone	-0.379 [-0.914,0.156]
Procurement	0.272 [-0.383,0.927]
t	-0.008 [-0.018,0.003]
t ²	0.002*** [0.001,0.003]
t ³	-0.000*** [-0.000,-0.000]
$\hat{\sigma}^2$	0.717

Results of a difference-in-differences regression of the change in the number of articles produced by a given journalist in a given month against the change in training status. The effects of training are differentiated according to the type of training received. Controls for a cubic polynomial of time adjust for trend effects. 95 percent confidence intervals are reported in brackets below the coefficient estimates. * denotes significance at the 95 percent level, and ** denotes significance at the 99 percent level. $\hat{\sigma}^2$ denotes the standard error of the regression – roughly interpreted as the average difference between the observed outcome and the prediction of the regression. All standard errors are clustered by journalist.

it amounts to roughly a 0.75 standard deviation increase in the change in the number of articles produced, relative to those who were not trained. The coefficient on the Lake Zone batch training is also suggestive of a positive effect of training on the number of articles produced. It is positive and significant at a 90 percent level of confidence.

The Powers training session was the most intensive of those offered by PACT. This training session focused on investigative methods to be applied to a broad variety of subjects. In addition to several days of seminars (which were offered in all training sessions), the Powers session provided travel grants to participants which were used to help complete a story which was previously being investigated. For instance, one journalist used her grant to investigate allegations that a local government was improperly allocating land rights. Participants would then bring the stories developed with the help of these grants back for criticism by the rest of the group.⁸⁵ Other training sessions did not involve this hands-on training process. Perhaps the hands-on and intensive nature of the Powers session explains why this round of training appears to have had a significant effect on the number of articles produced by participants; while other rounds did not.

However, as noted above, caution must be used in interpreting these results. Because we make multiple comparisons in this regression, we are likely to find positive associations between training and the change of the number of articles produced for at least a small number of training sessions. To be confident that these results reflect a true causal effect, we must ensure that they are robust across a number of alternative specifications. We examine the Powers and Lake Zone results in several alternative specifications below.

To test whether the Powers session truly had a positive effect, we compare those journalists who received the Powers training to *all other journalists who received training*. These results are reported in the first column of Table 8 below. We also compare the Powers journalists to all other journalists – both those who received training and those who did not. And, we allow the effects of training to vary over time.⁸⁶ These results are reported in

⁸⁵Some participants in earlier rounds of training were also offered the opportunity to apply for travel grants and to participate in the group criticism of resultant stories – i.e., some journalists who had already received training were invited to participate in the latter portion of the Powers session activities. However, participation in the travel grant and group criticism activities was only universal for the Powers session trainees. Our coding only captures those who were officially part of the full Powers session during this period. If participation in the travel grant and criticism activities boosted the output of other journalists, our estimates above underestimate the effect of the Powers session on journalist output.

⁸⁶We examine the effects of the Powers training over time both by including a cubic polynomial of the number of months since training, as reported in Table 8, and by adding indicator variables for each month following training (Laporte and Windmeijer, 2005, following). The results of these alternative strategies are substantively identical.

column two of Table 8 below.

In both instances, the estimated effect of the Powers training remains robust. It is positive, significant at the 99 percent level, and substantively large. Nor does this effect appear to diminish over time.

To ease interpretation, we present a simulation based on the results reported in column 2 of Table 8. Figure 6 plots the difference between the change in the number of articles produced by a journalist who received the Powers training and one who did not. The solid line represents the expected difference; while the dashed lines represent 95 percent confidence intervals. The number of months since training is represented along the x-axis (training takes place when this is equal to zero). The difference in the rate of change in the number of articles written in plotted on the y-axis.

As is evident from the figure, the Powers training session boosts the rate of change in the number of articles published by trained authors in the month that training is provided. In subsequent months, the rate of change is identical for those who did and did not receive the Powers training. This one-time shift in the rate of change in the number of articles produced corresponds to a permanent shift in the absolute number of articles produced per month. Those who were trained experience a one-month acceleration in the rate at which articles were produced, and no subsequent deceleration.

While the results regarding the Powers training session proved robust; those regarding the Lake Zone training did not. As robustness checks, we compare those that received the Lake Zone batch training to all other trained journalists; to all other trained journalists except those trained in the Powers session, and to all journalists; both trained and untrained. These results are reported in columns 1-3 of Table A.3 in Appendix A.3, respectively. In column 3, we also allow the effect of the Lake Zone training to vary over time. While the estimated coefficient on the Lake Zone training is consistently positive; it is not significant in any of these models.

THE EFFECT OF TRAINING ON THE QUALITY OF ARTICLES PUBLISHED

In this section, we attempt to assess whether, on average, the quality of a given journalist's writings improve following training. As described above, we rely on a difference-in-differences estimator to compare the rate of change in quality of trained journalists' articles to those of untrained journalists. The coefficient estimates from our regression should be interpreted as reflecting the difference in the average rate of change in the quality of trained journalists' articles relative to that of untrained journalists' articles.

Results from this specification are reported in column 1 of Table 9. Surprisingly, the

Table 8: Effect of Powers Session

	Δ No. Articles	
Δ training	0.006	
	[-0.129,0.142]	
Δ Powers	0.545**	0.551**
	[0.167,0.922]	[0.198,0.904]
t	-0.008	-0.008
	[-0.018,0.003]	[-0.018,0.003]
t ²	0.002***	0.002***
	[0.001,0.003]	[0.001,0.003]
t ³	-0.000***	-0.000***
	[-0.000,-0.000]	[-0.000,-0.000]
Months since Powers		-0.024
		[-0.055,0.007]
Months since Powers ²		0.003
		[-0.003,0.010]
Months since Powers ³		-0.000
		[-0.000,0.000]
$\hat{\sigma}^2$	0.717	0.717

Results of a difference-in-differences regression of the change in the number of articles produced by a given journalist in a given month against the change in training status. The coefficient on Δ Powers denotes the additional effect of the Powers Session relative to all other forms of training in the first model. It denotes the effect of the Powers Session relative to all other journalists in the second model. Controls for a cubic polynomial of time adjust for trend effects. Lagged effects of training are controlled for, in the second model, through controls for a cubic polynomial of the number of months since training. 95 percent confidence intervals are reported in brackets below the coefficient estimates. * denotes significance at the 95 percent level, and ** denotes significance at the 99 percent level. $\hat{\sigma}^2$ denotes the standard error of the regression – roughly interpreted as the average difference between the observed outcome and the prediction of the regression. All standard errors are clustered by journalist.

Table 9: Estimated Effect of Training on Article Quality

	Baseline Model	Partisan Model
Training	-0.287 [-0.592,0.017]	-0.058 [-0.586,0.470]
Pro Opposition Paper		0.562*** [0.247,0.876]
Pro Gov't Paper		-0.075 [-0.740,0.589]
Pro Opposition Paper * Training		-0.282 [-0.710,0.145]
Pro Gov't Paper * Training		-0.232 [-0.650,0.186]
Author Fixed-Effects	✓	✓
Time Fixed-Effects	✓	✓
$\hat{\sigma}^2$	0.499	0.493

coefficient on the training indicator is negative and significant at the 90 percent level. The point estimate of this coefficient indicates that training *reduces* the average quality of an article produced by slightly more than 0.33 standard deviations.

This result is quite puzzling. It is unlikely to stem from measurement error in the construction of the quality index, as difficulties in measurement would tend to reduce the significance of any finding.⁸⁷ Measurement error in the quality index would not bias coefficient estimates, provided we are as likely to overestimate as underestimate article quality.

One possibility is that our measure of article quality reflects different elements of quality than those emphasized in PACT's training. Our measure weights factors such as the quality of the writing and the ability to place a given story in a broader context; whereas, much of PACT's training emphasized investigative methods.⁸⁸ To the extent that inves-

⁸⁷Moreover, similar results are obtained if the raw 'overall quality' indicators coded by the enumerators are used in place of the quality index constructed above.

⁸⁸This difference in emphasis is, to some extent, unavoidable. The quality of an investigation is very difficult to judge from reading a finished article; while the quality of writing and presentation are more

tigative training encourages journalists to focus more heavily on the specifics of a given investigation, it may tend to reduce an emphasis on broader context.⁸⁹ However, lower scores on the quality index also imply that a given article is less likely to cite a large number of sources.⁹⁰ This would seem to suggest that lower scores also correlate with lower investigative quality.

It may also be the case that more and better investigations by trained journalists affect the quality of the output of untrained journalists. Trained journalists may uncover stories through their investigative work, which are then picked up by other media outlets.⁹¹ If this is the case, our estimates would understate the full impact of training. Training would boost the quality of *all* articles. There is some evidence of this – we find that article quality is improving from one quarter to the next, sometimes significantly so. But, such spillovers cannot fully explain the negative finding reported in Table 9. Spillover effects would tend to shift our estimates towards zero – a null effect. They would not tend to produce estimates of detrimental (negative) effects.

In an effort to further parse out the puzzling effects of PACT's training activities, we estimate an alternative model adjusting for the partisan nature of the newspapers in which journalists publish. While many politically neutral media exist in Tanzania; a number of outlets are known for their strong partisan (and sometimes religious) affiliations. Since the CCM is the dominant (and governing) party, these affiliations can generally be described as either favoring or opposing the CCM (the government). It may be the case that trained journalists that publish their articles in particularly partisan papers tend to use their investigative expertise primarily to attack their partisan opponents, rather than to develop broader lessons regarding corruption and good governance in Tanzania. The effect of training may therefore differ significantly if provided to the partisan, as opposed to the neutral press. We rely on the assessment of the researchers at EcomResearch to define newspaper partisanship. The partisan affiliations of the newspapers included in this sample are reported in Table A.4 in Appendix A.3.

The results of this model are reported in column 2 of Table 9. The coefficient on the readily identifiable. Our measure of quality does attempt to capture investigative quality through questions regarding the number of citations. But this measure is likely to be imperfect.

⁸⁹If regressions are run on the 'context' indicator directly, the coefficient on training is consistently negative for all enumerators. However these results are not significant at conventional levels.

⁹⁰Regressions run directly on the 'number of sources' indicator are not of a consistent sign across reviewers and are not significant.

⁹¹More formally, it may be the case that the Stable Unit Treatment Value Assumptions fail to hold in this instance.

training indicator, which now reflects the effect of training on the quality index scores of articles in neutral publications, is close to zero and far from significant. The coefficients on the interactions between training and the partisan affiliation of newspapers, however, are both highly negative. Indeed, the effect of training on publications in pro-government newspapers (which is equal to the sum of the coefficient on the training term and the interaction) is negative and marginally significant at the 90 percent level. The effect of training on pro-opposition publications is also estimated to be negative, though it is not significant at conventional levels. These results suggest that papers affiliated with political groups tend to use any advantages gleaned from training to score political points, rather than casting light on broader issues of corruption and governance.

Interestingly, the coefficient on the pro-opposition dummy is positive, large, and significant at the 99.9 percent level. This indicates that journalists (both trained and untrained) who publish in both neutral and opposition-leaning papers, tend to publish much higher quality articles in the opposition papers.

Sustainability

The training program established by PACT did not continue after the TCP closed. As we do not find a positive effect from training on either the quantity of articles produced or the quality of those articles for the average participant, we cannot speak of sustained gains from training in the aggregate. We do find that the Powers training session boosted the quantity of articles produced by participants, and that this effect did not appear to diminish over time. Our estimates also indicate an upward trend in article quality over time. This suggests sustained improvement in the quality of journalism in Tanzania – though we cannot definitively attribute these results to the training provided by PACT.

Summary of Results

As noted above, quantifying the impact of investigative training on the quantity and quality of articles produced is a difficult task. In particular, our estimates suffer from our inability to identify the source of anonymous articles. And our comparisons of trained to untrained journalists may be problematic if there are spillover effects from training. Nonetheless, our results provide valuable insight into the successes and failures of the journalist training program. A summary of our results is reported below.

Effects of Training on the Number of Articles Produced:

- Our estimates cannot support the conclusion that training led to an increase in the number of articles produced by trained journalists, on average.
- However, the training provided during the Jonathan Power’s training session *is* associated with a large and sustained increase in the number of articles produced by trained journalists.

Effects of Training on the Quality of Articles Produced:

- Our estimates indicate that training may have marginally *reduced* the quality of the average article produced by trained authors.
- Investigative training appeared to have a negative effect on articles produced in partisan newspapers – particularly in pro-government papers. It is likely that this effect had to do with a focus on partisan investigations that did not cast light on larger issues relevant to Tanzania.
- Estimates of the effects of training on the quality of publications in neutral papers are close to zero.
- However, our estimates may underestimate the positive effects of training to the extent that training had spillover effects on the quality of articles. Such spillover effects may exist if untrained journalists develop improved stories as a result of the investigative activities of trained journalists.

3.4.2 Public Expenditure Tracking Surveys

Background

The expansion of local public expenditure tracking was one of the central outputs to be achieved from the Tanzania TCP. Specifically, the TCP called for “An increase in the number of local government authorities participating in public expenditure tracking from 12 to 60.”⁹² To this end, Campaign for Good Governance (CGG) was awarded \$200,000 to provide Public Expenditure Tracking Survey (PETS) training in 12 districts and PACT-Tanzania (PACT) was awarded \$2,615,000 to provide PETS training in 60 districts.⁹³

PETS councils were established in 77 of Tanzania’s districts to conduct public expenditure tracking. These councils were composed of local citizens who would review local

⁹²The Strategic Objective Grant Agreement.

⁹³The grant awarded to PACT was also intended to help strengthen civil society organizations with whom PACT partnered to provide PETS training.

governments' budgetary decisions. This review process was intended both to note any accounting irregularities and to discover any discrepancies between local governments' expenditures and the services voted upon in village-level community meetings. Discrepancies were to be made public through a system of centrally located notice boards.

The methods used to carry out training differed between the two implementers. CGG provided training directly to the 12 districts in which it was involved. It established PETS councils at the district-level and at the village-level for all villages within a given district. PACT incorporated PETS training under its Building Organizational Networks for Good Governance and Advocacy (BONGA) initiative – which had been supported by USAID since 2005. It relied on a system of civil society organizations (CSOs) to provide training. These CSOs could choose to deliver PETS training in districts in which they were already active. Individual CSOs determined whether PETS councils would operate at the village- and district-levels, or only at the district-level.

Methodology

To quantify the effect of PETS training, we rely on local government audit reports produced annually by the Controller and Auditor General (CAG) of Tanzania.⁹⁴ Local government authorities must submit financial statements to the CAG's office annually. Members of the CAG's staff then perform a financial audit of these documents, noting discrepancies and missing funds. Each local government authority is also assigned one of three grades, based on the quality of its record-keeping.

In our evaluation, we examine the percentage of expenditures that are questioned by the CAG (both overexpenditures and underexpenditures). We also examine the opinions of the auditors regarding the quality of financial reporting. These opinions may take one of three values: unqualified, qualified, and clean.⁹⁵ The data run from FY 1998/9-2008/9 – though, data are not available for FY 2007/8.

Our analysis compares changes in the percentage of questioned revenues and in auditor opinions between districts that received PETS training and those that did not. This methodology is a difference-in-differences design. The difference-in-differences model is advantageous in this case because districts were not selected at random for training.

⁹⁴Article 143 of the Constitution of the United Republic of Tanzania and Section 45 of the Local Government Finances Act of 1982 require that local government authorities submit annual financial reports to the office of the CAG.

⁹⁵The classifications for auditor opinions retained this ordinal character; though the labeling of the three opinions changed in FY 2008/9.

Rather CGG and the various CSO organizations supported by PACT selected which districts would receive training. The difference-in-differences design controls for all district characteristics that remain fixed over time. Differences in district characteristics are only relevant insofar as they effect the *rate of change* in auditor opinions or questioned revenue.

There does remain a danger, however, that district characteristics affect this rate of change. This is particularly true due to the nature of the measures we employ. Districts given ‘clean’ ratings by the CAG cannot improve on these ratings over time. And the percentage of revenues that are questioned cannot rise above 100 percent or fall below 0 percent. The level of these terms may thus effect their rate of change. If PETS training is therefore assigned in a manner that correlates with background variables, our measures may be biased.

To address this possibility, we employ a propensity score matching algorithm. Propensity score matching ensures that our estimates compare trained districts with similar untrained districts, where similarity is assessed according to covariates measured prior to training. More precisely, we estimate the probability that a given district received PETS training based on a full range of variables drawn from the CAG data measured between FY 1998/9 and FY 2006/7. Observations are then weighted such that untrained districts that are similar to trained districts are given greater weight than dissimilar districts. 20 untrained districts are dropped from our estimates. Matching is performed using the genetic algorithm of Diamond and Sekhon (2006) run from the MatchIt package in R 2.7.1.

There also exists a deeper danger – the information in the CAG reports may be flawed. These audits are based on information made available by local government authorities who may have an incentive to conceal financial mismanagement. To the extent that local authorities are sophisticated in manipulating their records, there may exist budget mismanagement that does not appear in the CAG data. If the level of misreporting is constant over time, it is unlikely to affect our results. If, however, it changes over time – our results may be confounded by these measurement errors.

We also perform a qualitative assessment of the PETS training based on interviews with PETS council chairs. 15 PETS council chairs were interviewed in August and September 2010. Questions examined whether the PETS councils continued to function, how often they meet, the qualifications and numbers of their members, and the types of activities they focus on. The answers to these questions both help us to assess the sustainability of the PETS training program and to examine possible effects that our quantitative estimates might not capture.

Outputs

A total of 77 districts received PETS training, as outlined in the ‘Background’ section.

Outcomes

QUANTITATIVE ESTIMATES

To assess the effectiveness of the PETS training program, we compare changes in the percentage of questioned expenditures in districts where PETS training was administered to districts where it was not. We also examine whether auditor opinions were more likely to improve in districts that received training than in districts that did not.⁹⁶

We examine changes in the questioned expenditures between FY 2006/7 and FY 2008/9. Unfortunately, because the CAG changed the method in which data is recorded in FY 2007/8, these data are not available for that year. On average, the percentage of questioned expenditures increased by 4.3 percent over that period. Though the distribution of the change in questioned expenditures is right skewed – the average district saw an increase of 2 percent in questioned expenditures.

The average district saw no change in auditor scores between FY 2006/7 and 2007/8. 14.8 percent of districts saw an improvement in auditor scores that year. 33.9 percent of districts saw a decline.

The results of our difference-in-differences comparison of the percentage of questioned expenditures is reported in Table 10. The coefficient on the PETS training term in Model 1 reports the average difference in changes in questioned expenditures between all districts that received PETS training and all districts that did not. The coefficients in this table represent the difference in the average percentage point change in questioned revenues between districts that received PETS training and those that did not. In Model 2, we differentiate between the effects of training provided by PACT and that provided by CGG. In this instance, the average difference between CGG trained districts and untrained districts is given by the sum of the coefficients on PETS training and CGG training.

In no instance is PETS training significantly associated with above average improvement on the percentage of questioned expenditures. Point estimates indicate that districts that received PETS training worsen their percentage of questioned expenditures by 0.9 percentage points more than untrained districts – though this effect is somewhat smaller in CGG trained districts. These results do not significantly differ from zero.

⁹⁶The outcome variable in this instance is trichotomous, assuming values of $-1, 0, 1$. A score of -1 denotes a worsening auditor score, a score of 0 indicates no change in score, and a score of 1 indicates an improvement in score. The empirical model is an ordered probit regression.

Table 10: Change in the Percent of Questioned Expenditures Associated with PETS Training

	Model 1	Model 2
PETS training	0.009 [-0.036,0.053]	0.009 [-0.037,0.056]
CGG training		-0.004 [-0.084,0.076]
$\hat{\sigma}^2$	0.120	0.121

Table 11: Change in Auditor Opinions Associated with PETS Training

	Model 1	Model 2
PETS training	-0.109 [-0.517,0.298]	-0.076 [-0.509,0.357]
CGG training		0.222 [-0.515,0.960]

Table 11 reports similar results for the ordered probit model examining changes in auditor scores. The results of this model indicate that PACT trained districts perform slightly worse on this measure than do untrained districts; while CGG trained districts perform slightly better. However, no difference is statistically significant.

As noted above, these comparisons may suffer from the fact that assignment to PETS training is non-random. Districts that receive PETS training may differ systematically from those that do not.

To adjust for this possibility, we employ a propensity score matching technique. We estimate the probability that each district in our sample is assigned to PETS training based upon average budgets *per capita*, budget totals, development aid, grants, other revenues, personnel expenditures, revenues, questioned revenues, questioned expenditures, and median auditor opinions over the FY 1998/9-2006/7 period. We then compare districts based on the estimated probability that each district will be assigned to PETS training. Untrained districts that are not sufficiently similar to any trained districts are pruned from the dataset. After matching we are left with a sample of 58 trained districts and 30 untrained districts. Matching diagnostics and results are reported in Appendix A.4. These results are qualita-

tively similar to those obtained from the full sample without propensity score matching.

It is possible that PETS councils do improve their district's budgeting procedures, but that our estimates – due to random error – fail to detect these effects. That is, it is possible that our results are the product of Type II error – we fail to reject the null hypothesis that PETS districts have no effect even though the null hypothesis is not true. It is therefore worth considering how large an effect would need to be present for us to find an effect.

To achieve statistical significance, PETS councils would need to reduce the percentage of questioned revenues by roughly 5 percentage points relative to districts that did not receive PETS training. That is, rather than a coefficient of 0.009, our regression results would need to report a coefficient of -0.044. The average district reported increasing questioned expenditures of 4 percentage points between 2007 and 2009 – thus, for PETS to have a significant effect, PETS trained districts would need to report no change.

An alternative calculation looks at the *statistical power* of our results. Statistical power refers to the probability that a Type II error is committed. A power test asks how large a true effect would need to be for the probability of a null finding to be sufficiently small. In this instance, PETS training would need to reduce questioned expenditures by 8 percentage points if the probability of Type II error were only 5 percent. That is, if PETS training reduce questioned expenditures by 8 percentage points, the probability that we find a null effect would only be 5 percent. This effect is quite large relative to the variation in the data.

It is therefore quite possible that PETS activities are actively performing their intended tasks, but that these are insufficient to have an appreciable effect on our empirical results. This may be because our tests lack sufficient power. Alternatively, it may be that the PETS councils lack sufficient resources to police the full set of expenditures by the district governments – and thus the effects of PETS activities are quite small relative to the total budget. In surveys of PETS council chairs, most report uncovering two to three instances of inappropriate accounting over the period since they began operations. While we lack data on the size of these discoveries; it seems unlikely that the number of irregularities uncovered is sufficient to cause a substantial decline in the percentage of questioned expenditures. Finally, it may be the case that PETS councils detect irregularities, but that these remain uncorrected. Such a situation might occur if the errors detected by PETS councils are not actively prosecuted. One PETS council member noted, in a conversation with the authors, that violations uncovered by her council frequently went unpunished.

QUALITATIVE RESULTS

To better assess the effects of PETS activities that may not directly translate into differences in government accounts, we conduct a series of semi-structured interviews with PETS council chairs. We randomly select PETS councils for interviews from the list of districts that receive PETS training, to ensure that our sample is representative. A list of councils participating in the interview process is contained in Table A.8 in appendix A.4. Interviews were conducted by EcomResearch in August and September of 2010.

The interviews cover matters related to the operation of a given PETS council and its performance in uncovering instances of misappropriation. The answers to these questions may provide anecdotal evidence of PETS activities that fail to translate into observable outcomes in our quantitative data and may reveal the sustainability or lack thereof of the PETS intervention. Appendix C contains a replication of the questionnaire provided to PETS council chairs.

Despite the lack of quantitative evidence of any effect of PETS training on local government budgets, the qualitative interviews indicate that the PETS councils remain active. 13 of the 15 councils interviewed meet monthly; while the remaining two meet quarterly. No PETS council reports any problem of attendance – indeed the 15 councils report an average 94 percent attendance rate for council meetings.

PETS councils also broadly appear well-suited to examine public finances. While most council members have a primary education; most councils contain at least one member who has experience in government or private sector accounting. Only one council lacks any member with financial experience. Most also report consulting with members of the public on a monthly basis.

However, the PETS councils' records of uncovering and reporting local government misappropriation are somewhat less encouraging. 12 of the 15 PETS councils interviewed report uncovering at least one instance of financial mismanagement – an average of 1.85 cases were uncovered by each PETS council since beginning operations. Moreover, not all instances of mismanagement were publicized. The relatively small number of cases of mismanagement uncovered by the PETS councils is consistent with the insignificant effects of PETS training uncovered in the quantitative results above.

The small number of cases may reflect the deterrent effect of the PETS councils. Government officials may exercise greater care in financial matters given the danger that mismanagement will be uncovered by the councils. Indeed, most PETS council chairs report that local governments are exercising such care. But, it seems unlikely that dramatic changes in government behavior would not be reflected in the budget audits conducted by

the CAG.

PETS council chairs also report greater public involvement in the budget process. And citizens, they claim, are now better informed of budgeting matters. This claim is certainly plausible, given the role of the PETS councils in disseminating budget information. However, without extensive surveys, it is impossible to confirm through quantitative analysis.

Sustainability

Qualitative fieldwork reveals that the PETS councils remain operational. Moreover, the Prime Minister's Office of Regional Administration and Local Government is promoting the expansion of PETS councils and has adopted the methodologies promoted during the TCP for this purpose. It therefore appears that the PETS intervention proved sustainable.

Summary of Results

Our findings regarding the PETS intervention are as follows:

- The program was implemented as intended. Indeed, the number of districts trained by PACT exceeded that planned for.
- However, we are unable to detect an effect of PETS council training on either the percentage of questioned expenditures or on auditor opinions of district-level government finances.
- Qualitative fieldwork reveals that nearly all councils interviewed report uncovering instances of improper accounting. Though the number of instances reported over the lifetime of the council is small.
- Qualitative fieldwork reveals that PETS councils continue their activities. The majority continue to meet on a monthly basis, and a large minority meet on a quarterly basis.

3.4.3 Public Awareness

Background

In April 2007, the Tanzanian Parliament passed the Prevention and Combating of Corruption Act. The central aim of the public awareness component of the TCP was to increase the public's knowledge of this Act and to facilitate the public's participation in

anti-corruption activities. Public participation was felt to be vital to the success of anti-corruption activities and to the fostering of anti-corruption norms of behavior.

To facilitate public awareness of the Prevention and Combating of Corruption Act, NOLA was awarded a grant of \$322,740.

To this end, the MCC contracted NOLA to conduct a public awareness campaign. This campaign consisted of the following activities:

- A series of radio and television spots.
- The placement of billboards and posters in public places.
- Public debates regarding corruption involving women, the youth, and the disabled.
- A series of workshops for members of the media regarding the particulars of the Preventing and Combating of Corruption Act.
- The publishing of pamphlets/primers on the Prevention and Combating of Corruption Act for public dissemination.

Methodology

Unfortunately, no attempts were made to survey public attitudes towards corruption or knowledge of the Prevention and Combating of Corruption Act prior to the implementation of the public awareness program. We are thus unable to quantify changes in public opinion and knowledge over time. Moreover, even were such information available, it would be difficult to attribute any changes in these measures to the public awareness campaign conducted by NOLA.

As we are unable to clearly identify the causal effect of the media awareness campaign, we conducted a qualitative review of NOLA's activities. We compare the activities that were carried out to those initially planned under NOLA's contract. And we consider the likely effect of these activities in light of the academic literature on corruption, public opinion, and media campaigns.

Outputs

Over the course of the TCP, NOLA achieved all outputs.⁹⁷

⁹⁷Participants in NOLA's interventions were unable to meet with the authors during the evaluation process. This assessment is based on Quarterly Reports, Macro International's performance audits, and interviews with Macro staff.

Outcomes

The aims of NOLA's public awareness program were sensible and in keeping with the overall goals of the TCP. Public involvement is likely to prove critical to successfully combating corruption. Authorities often rely on reports from members of the public to launch investigations. Constituents must hold politicians to account for corrupt activities for the government to have a strong incentive to crack down on corruption. And there exists substantial empirical evidence of the force of a 'culture of corruption' (see, for instance, Fisman and Miguel, 2007). The prevailing perception that corruption is a normal activity – and the expectation that others will also engage in corruption or will ignore corruption if observed – helps to reinforce corrupt patterns of behavior.

There is also reason to think that media campaigns may be effective in changing norms and behavior. Paluck and Green (2009) find that mass media interventions (radio plays) were effective in altering norms regarding dissent and debate in Rwanda. It is less clear whether such media campaigns are sufficient to convey the complexities of anti-corruption legislation – or how detailed such information would need to be to have the desired effect on behavior.

However, the length and intensity of the media campaign may have proved insufficient to substantially alter patterns of behavior amongst members of the public. NOLA staff expressed concern that the relatively short duration of the TCP hindered efforts of both the media training and the public outreach activities.⁹⁸ Exposure to the media campaign over the May 2007-May 2008 period would have to be quite high to have a discernable effect on behavior. Any effect would be concentrated on those members of the public most exposed to the media (for more on heterogeneous effects of media exposure see Barabas and Jerit, 2010).

Several concerns were also voiced over the quality of the materials used in the public awareness campaign – particularly with one of the television spots developed. One television spot seemed to target citizens of a foreign country as agents of corruption. This spot was eventually dropped.

However, the media training activities were positively viewed by Program participants. Participants interviewed for the Macro International Mid-Term and Final Performance Audits expressed positive opinions regarding the content of the media training. Moreover, NOLA faced greater demand for training than first anticipated, such that a larger number of journalists and editors were trained under this intervention than initially anticipated in

⁹⁸Final Performance Audit MCA/Threshold Program Tanzania. Macro International. November 2008.

NOLA's contract.

Sustainability

At the close of the TCP, NOLA was in the process of drafting a three year strategic plan. This plan allowed for public awareness activities regarding the Prevention and Combating of Corruption Act to continue, provided that stakeholder interest remained strong. Unfortunately, however, meetings with NOLA staff could not be scheduled and we cannot confirm whether activities are ongoing.

However, we can confirm that the PCCB continued to distribute several of the brochures developed by NOLA after the close of the TCP.

Activities aimed at directly informing members of the public are less likely to have a sustained impact. Much of the literature on political behavior finds that informational interventions tend to have a relatively short-term effect on beliefs and attitudes.

Summary of Results

Unfortunately, we are unable to quantify the effect of NOLA's activities on outcomes of interest – namely the beliefs and attitudes of members of the public. The conclusions of our qualitative investigation into these outcomes are as follows:

- Efforts aimed at training members of the media were administered to the desired audience and anecdotal accounts suggest they were effective.
- While the aims of the public outreach campaign were appropriate; participants raised concerns over the duration and scope of this campaign. We lack adequate information to determine whether these concerns were merited. But the academic literature on media effects finds that the duration and scope of informational campaigns are critical to their effectiveness.

3.4.4 Wildlife Management Areas

Background

A contract of \$100,000 (of which \$50,000 had been distributed by November 30, 2008) was awarded to the African Wildlife Foundation (AWF) to develop business and community outreach plans for two Wildlife Management Areas (WMAs). Wildlife Management Areas (WMAs) are organizations created by the Government of Tanzania to devolve control of natural resources and land management to local government. Plans to establish

WMAs in Tanzania date to the Village Land Act of 1999. WMAs are formed by an association of several villages that border on national parks. These villages may choose to form a community-based organization (CBO) that will manage their joint village lands. Representatives are elected to the CBO from each village in the WMA. Each village must select one male and one female representative. Once a CBO is formed and registered, it must develop a land-use plan for the disposition of village lands. Land-use plans must be approved by the Wildlife Division of the Ministry for Natural Resources and Tourism and must meet environmental standards. To aid in the process of registering the CBO and approving the land-use plan, each WMA is required to have an NGO partner organization. AWF, the contractor selected for this intervention, is one such NGO.

At the time the Tanzania TCP was implemented, 16 WMAs had recently begun operations. To support the activities of some of the newly-operational WMAs, USAID modified the Tanzania TCP in late 2007 to include the creation of two integrity pacts between WMAs and private sector firms involved in the game and tourism sectors. The WMA program began to show momentum at this time, owing in part to a changes in the upper levels of the Wildlife Division of the Ministry for Natural Resources and Tourism. USAID saw an opportunity to capitalize on this progress to advance the aims of the TCP.

Under its contract, AWF was to accomplish the following:

- Conduct an economic analysis for potentially attracting private sector investment in two WMAs (Burunge and Enduimet),
- Develop a business plan for both of these WMAs,
- Develop a best practices toolkit for WMAs,
- Help execute any contracts between the WMAs and private sector investors.

The purpose of these activities was to attract private sector investment to each WMA in a manner that would ensure maximal public oversight and public benefit. These activities were seen as part of a larger effort to reduce corruption in the tourism and game sector.

Methodology

In light of the small number (2) of WMAs involved in this intervention, and given the difficulties of collecting adequate data on measures such as quality of life for local residents and the level of corruption in the tourism and game sector, we confine ourselves to a qualitative investigation of this intervention. We focus on the AWF's ability to deliver the

outputs listed under its contract. And we discuss the likely outcomes and sustainability of these efforts, taking into account both materials acquired over the course of our evaluation.

Outputs

The AWF delivered all intended outputs under this intervention.

Outcomes

This intervention bears a logical relation to corruption. Before the creation of the WMAs, the AWF reports that local officials would often accept bribes from hunting and tourist operators in exchange for access to village lands. The WMA structure helps to reduce the opportunities for such corruption in land management. WMAs restructure the procedures by which land use is governed, granting greater influence to local representatives. Perhaps more importantly, local citizens are given a financial stake in holding their representatives to account – as revenues from tourism of WMA lands are devoted to local infrastructure development (and met with matching funds by the government).

The provision of a best practices toolkit and business plan fits a logical need for WMAs to be successful. These toolkits place local representatives in a better negotiating position vis-à-vis potential investors. And, they help to inform local residents of the positions their representative *should* be taking. Indeed, after AWF began developing business plans for its WMAs, other partner agencies (in other WMAs) also began to adopt the practice.⁹⁹

Following this intervention, the Burunge WMA established a partnership with a gaming and tourism operator. According to the AWF, the percentage of revenues from tourism that went to the villagers jumped dramatically – more than threefold. The Wildlife Division of the Ministry of Natural Resources and Tourism reports that the Burunge WMA saw revenues of \$72,000 in 2007; while the Enduimet WMA saw revenues of \$78,000. Unfortunately, we do not have earlier numbers to compare these figures to and thus corroborate AWF reports.

Sustainability

Many of the benefits of these WMAs rely on institutional innovations that are likely to be sustained over time. Institutions are, in general, slow to change – and the WMAs enjoy the support of the Tanzanian government. Moreover, AWF continues to play an active role

⁹⁹Interview with officials at the Wildlife Division of the Ministry of Natural Resources and Tourism. April 2010.

in advising local CBOs in both WMAs, which should help to ensure that the benefits of the Business Plans and Best Practices Toolkit continue to be felt.

AWF currently receives support from USAID. Indeed, AWF and USAID are jointly supporting two new WMAs in Kilindi and Yaedachini.

We are also encouraged to notice that the WMA program has expanded from 16 to 33 WMAs. This suggests that WMAs continue to enjoy government support and are widely seen as a successful innovation. Moreover, as noted above, other WMAs have come to adopt the practice of drafting business plans to partner with private sector investors – following the model adopted by the AWF and supported by the Tanzania TCP.

Summary of Results

We thus note the following outcomes with regard to the WMA component of the Tanzania TCP:

- Business plans and a Best Practices Toolkit were developed for both the Burunge and Enduimet WMAs.
- The Burunge WMA secured a private sector investor in the wildlife and game sector, and – according to AWF – substantially increased revenues as a result.
- The list of WMAs is expanding and other WMAs have also come to adopt the process of drafting business plans to attract private investment.

4 Conclusions and Lessons Learned

Our results indicate that the vast majority of TCP activities were completed as intended. The lone exception was the MP training program which was curtailed early. USAID also showed flexibility in amending planned activities to include new interventions where opportunities arose – e.g., the DPP training sessions and the WMA business plans – and in abandoning interventions viewed as extraneous to the mission of the TCP – e.g., training the Human Rights Commission.

Our quantitative evaluations are generally unable to uncover evidence for an effect of TCP interventions on corruption. The striking exception is the system of audits administered by Kilimanjaro International and the PPRA – which produced a statistically significant and substantively large improvement in procuring entity compliance with procurement

regulations. We also find evidence that one of the ten (9 tested) journalist training sessions increased the quantity of articles produced by participants – though this effect was absent in other training sessions. Twenty eight journalists – of the 312 journalists trained – took part in this session.¹⁰⁰

In general, participants expressed concern that many of the training sessions offered under the TCP were too short or infrequent to achieve their desired effects. The most successful training – according to both quantitative and qualitative results – were more intensive and involved more practical activities – e.g., the travel grants provided to journalists in the Powers’ session.

The TCP also successfully created a number of institutions that appear to be sustainable – the PETS councils, the FIU, and a legal aid secretariat. We are unable to adequately assess the effectiveness of many of these institutions through quantitative methods. It is encouraging, however, that many such institutions continue to perform their intended functions, many of which bear a logical relationship to anti-corruption efforts.

4.1 Lessons Learned

- **Embedded training exercises are more effective than short-term training seminars.** Much of the Tanzania TCP focused on training civil servants or members of civil society. Training generally took one of two forms: short term seminars (generally of one week or less), or more intensive training, often involving side-by-side work with embedded trainers. Examples of the latter form of training include the auditing process conducted jointly by the PPRA and Kilimanjaro International, the corruption investigations conducted jointly by the PCCB and NYC DoI, and the provision of travel grants to participants in one of the journalist training sessions. Quantitative investigations of the auditing process and of the Powers training session find that both had significant impacts on, respectively, compliance with procurement regulation and the production of articles related to corruption. Qualitative results similarly suggest that the joint NYC DoI/PCCB investigations successfully concluded investigations into corruption and facilitated the transfer of skills to the PCCB. By contrast, we were unable to detect any effect of short training seminars offered either to journalists or to the PCCB.

It should be cautioned that the difference in the effectiveness of training sessions

¹⁰⁰Participation in the travel grant program that this training session incorporated was more broad.

cannot be treated as causal. That is, we cannot say that replacing short-term seminars with more intensive exercises or embedded trainers would necessarily produce better results. The forms of training that were offered to Program participants were influenced by the willingness and ability of these participants to devote time and staff to training exercises. The differences in the effectiveness of training may, therefore, be attributable to selection effects. Regardless of the causal mechanism at work, however, these results do suggest that interventions that rely on short-term training seminars are less likely to be effective – perhaps because participants are unable or unwilling to devote greater resources to training activities or because such seminars are themselves ineffective.

- **The two year time frame may limit the effectiveness of some interventions.** Limiting the duration of the intervention to two years may have both advantages and drawbacks. Participants (notably USAID) cite the limited duration as increasing the urgency of TCP implementers in carrying out the implementation of the program. However, the two year period proved insufficient for some interventions to be completed, and may have hindered the effectiveness of others.

Unfortunately, we cannot readily attest to the benefits of a two year implementation period. It is certainly possible that the limited time frame enhanced participants' sense of urgency. However, urgency is a difficult concept to adequately measure, even informally. And it is still more difficult to assess the urgency of implementers two years after the close of the TCP. Moreover, there was no variation in the time frame set for implementers – the two year limitation applied to all. So we cannot say whether those faced with a more limited time frame acted in a more urgent manner.

The drawbacks to the two year implementation period are more readily visible. Several TCP components could not be completed on time and others may have proved more sustainable if implemented over a longer period. This claim most obviously pertains to the creation of the FIU. The TCP called for the passage of an anti-money-laundering law and the creation and staffing of an FIU in a two year time frame, such that the FIU would be able to begin operations by the close of the TCP. In practice, selection of a contractor to aid in the creation of the FIU did not take place until the end of the first year of the Program, the confirmation of an FIU commissioner engendered further delays, and – by the end of the TCP – the FIU was neither fully staffed nor fully operational. This shortage of staff persisted through April 2010. The short

time frame of the TCP also limited the duration of programs intended to promote public awareness of corruption. Several participants in NOLA's media training sessions expressed an interest in further training that could not be provided due to the short duration of the TCP. The two-year limit on TCP activities also limited the length of possible public awareness campaigns – and research indicates that the length and intensity of exposure to such messages have significant effects on their effectiveness (Barabas and Jerit, 2010).

At the least, these results suggest that, if a two year time frame is deemed desirable, only those interventions appropriate for such a time frame should be adopted. The creation of a wholly new regulatory entity – such as the FIU – is unlikely to be feasible in so short a period. Similarly, public awareness campaigns will likely require a longer period to have a sustainable effect on public attitudes towards corruption. Training programs or the creation of local entities – such as the PETS committees – are more likely to be practical in a short period.

- **The TCP often achieved successes by contracting implementers with established relationships with Tanzanian government agencies.** This lesson is most clearly in evidence for the interventions targeting the PCCB and the WMAs. The PCCB specifically asked USAID to contract a portion of the Rule of Law component of the TCP to NYC DoI. This request resulted directly from the long relationship between the NYC DoI. The trust between the two agencies was such that the PCCB proved willing to allow NYC DoI officials to be embedded in two prominent investigations currently underway. Similarly, USAID contracted the WMA intervention to the AWF – an organization long involved in the Burunge and Endiumet WMAs. As a result of this relationship, the AWF was already familiar with the operations of these two WMAs and was able to successfully complete a business plan for each. According to anecdotal evidence, the Burunge WMA was able to secure a private sector investor for, substantially boosting tourism related revenues by the conclusion of the TCP.

More precisely, these relationships contributed to program successes by building off successful work in the past. AWF had successfully guided the Burunge and Endiumet WMAs through the registration process and was already working on fostering development in these areas. The relationship between the PCCB and NYC DoI was, in part, founded by the strong personal relationship between the NYC DoI Deputy Director and the PCCB Director. Their past interactions were such that Director Hoseah

specifically requested the NYC DoI's involvement in TCP sponsored training.

These results suggest that the MCC would benefit by (1) contracting with organizations that have a proven track record of success in interactions with program partners, and (2) would likely benefit from heeding participant recommendations in the selection of Program implementers.

- **The flexibility of USAID to new opportunities allowed for several well-targeted interventions to be implemented.** This finding holds particularly true for the intervention targeting the DPP and the WMAs. At the time of the TCP, the DPP was transforming its role as part of the 'civilianization' of public prosecutions. As a result, the DPP needed a large number of new staff, many of whom required training. USAID was quick to respond to these circumstances and crafted a training program that covered all new prosecutors. Anecdotal accounts suggest that this training process was highly successful. USAID demonstrated similar flexibility in responding to progress in the creation of the WMAs. While the Tanzanian government had long planned to devolve wildlife management authority to these local bodies; substantial progress was realized in the creation of WMAs only after the TCP opened. USAID responded to these changes and was able to successfully support the creation of business plans in two WMAs.

The responsiveness of USAID to emerging opportunities is largely a testament to the abilities of its staff – particularly the Program Manager. It is likely that the responsiveness of the Tanzania TCP was attributable to the clear definition of its aims – namely the reduction of corruption. In the absence of a clear statement of purpose, there is greater room for misguided interventions that bear little relation to Program aims or to one another.¹⁰¹

It is also probable that a strong M&E effort is complementary to program flexibility. Ongoing evaluation efforts help ensure that wasteful or ineffective interventions may be recognized early and canceled or curtailed. Program flexibility helps to ensure that staff can respond to the M&E findings as they are made available.

- **Greater selectiveness was warranted to ensure that interventions targeted cor-**

¹⁰¹Indeed, even in the presence of clearly stated aims, several interventions under the Tanzania TCP lacked a clear logical relationship to the reduction of corruption. If aims were less clearly defined, such problems may have been exacerbated. A substantial literature in political science and economics finds that accountability is reduced as the number of dimensions across which performance is measured expands (see, for instance Holmstrom and Milgrom, 1991).

ruption. Several of the interventions undertaken as part of the TCP bore an indirect or questionable relation to the goal of combating corruption. The link between corruption and the expansion of the legal aid system was particularly tenuous. Some of the training sessions administered to the PPRA – for instance workshops in team building – similarly bear little relation to combating corruption.

- **Interventions that sought to create new institutions often proved more sustainable than capacity-building exercises.** A variety of new institutions were created as part of the TCP, notably: the PETS councils, a standardized auditing procedure for procuring entities, a legal aid secretariat, and the FIU. While we are unable to document an effect on corruption from several of these institutions; all remain in operation two years after the close of the TCP. By contrast, many programs aimed at building capacity – the training of journalists and other members of the media, the training provided by ICITAP and OPDAT to the PCCB – ended at the close of the TCP. While different forms of training may continue at the PCCB; ICITAP and OPDAT do not conduct these sessions. And PACT is not currently involved in journalist training. Institutions, therefore, appear to have a more persistent impact than do capacity building exercises.
- **Greater attention to M & E from the inception of the Program would like improve both the ability to detect the impact of interventions and may have prevented the implementation of interventions without a clear logical connection to the aims of the Program.** Assessment of the impact of the Tanzania TCP was frequently hindered by the limited availability of data and the difficulty of constructing empirical strategies to identify the effect of interventions after the fact. The vast majority of the data used in this report were collected for reasons unrelated to the TCP – for instance, the records of the PCCB and the local government audits used to evaluate the PETS program. As these data were not collected with the intention of measuring the effectiveness of TCP interventions, they may imperfectly reflect the effectiveness of these interventions. For instance, the local government audits may be too highly aggregated to capture the effect of PETS committees on local budgeting. Were M&E more thoroughly integrated from the inception of the Program, it is likely that indicators could more closely mirror the intended effects of interventions.

Had M&E been a greater consideration from the inception of the TCP, subsequent evaluation might also have been better able to assess the effectiveness of the Program

in producing its desired outcome – reductions in the level of corruption – rather than focusing on outputs – e.g., the behavior of anti-corruption agencies, or of trained journalists.

Problems of causal identification may have been eased though greater use of randomization of treatment (e.g., training) or in the timing of treatment. The TCP did not employ randomization in any of its interventions, leading to possible concerns that the effects of such interventions may be confounded by selection bias. Moreover, in many instances, interventions were administered once on a national scale. This design of interventions ensures that there is no variation in treatment across individuals or areas that allows evaluators to assess their efficacy. In some cases (e.g., the creation of the FIU), this design was unavoidable. However, interventions such as training or public awareness programs can be varied over time or geographic areas.

Finally, a greater focus on M&E would have necessitated careful consideration of the logic of all interventions, and may have prevented the implementation of interventions whose logic was questionable. For instance, we find that the logic of both the legal aid intervention and the training of MPs in procurement law bore a questionable logical link to the goal of limiting corruption. Early involvement of evaluators may have noted the weak logical links between these interventions and corruption and canceled these interventions. Weak logical links are particularly likely to be highlighted when attempting to design measures for evaluation.

References

- Alesina, Alberto and Guido Tabellini. 2007. “Bureaucrats or Politicians? Part I: A Single Policy Task.” *The American Economic Review* 97(1):169–179.
- Alt, James E. and David Dreyer Lassen. 2010. “Enforcement and Public Corruption: Evidence from US States.”
- Barabas, Jason and Jennifer Jerit. 2010. “Are Survey Experiments Externally Valid?” *American Political Science Review* 104:226–242.
- Becker, Gary S. and George J. Stigler. 1974. “Law Enforcement, Malfeasance, and Compensation of Enforcers.” *Journal of Legal Studies* 3(1):1–18.

- Besley, Timothy. 2006. *Principled Agents? The Political Economy of Good Government*. Oxford University Press.
- Besley, Timothy and Stephen Coate. 2003. "Elected Versus Appointed Regulators: Theory and Evidence." *Journal of the European Economic Association* 1(5):1176–1206.
- Carter, David B. and Curtis S. Signorino. 2007. "Back to the Future: Modeling Time Dependence in Binary Data." The Society for Political Methodology Working Paper.
- Clinton, Joshua and David E. Lewis. 2008. "Expert Opinion, Agency Characteristics, and Agency Preferences." *Political Analysis* 16(1):3–20.
- Clinton, Joshua, Simon Jackman and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *The American Political Science Review* 98(2):355–370.
- Diamond, Alexis and Jasjeet S. Sekhon. 2006. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies."
- Fisman, Raymond and Edward Miguel. 2007. "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets." *Journal of Political Economy* 115(6):1020–1048.
- Gelman, Andrew and Jennifer Hill. 2006. *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Analytical Methods in Social Science Cambridge University Press.
- Gordon, Sanford C. 2009. "Assessing Partisan Bias in Federal Public Corruption Prosecutions." *The American Political Science Review* 103(4):534–555.
- Holmstrom, Bengt and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics and Organization* 7:24–52.
- Huber, John D. and Charles R. Shipan. 2002. *Deliberate Discretion? The Institutional Foundations of Bureaucratic Autonomy*. Cambridge: Cambridge University Press.
- Humphreys, Macartan, William A. Masters and Martin E. Sandbu. 2006. "The Role of Leaders in Democratic Deliberations: Results from a Field Experiment in São Tomé and Príncipe." *World Politics* 58:583–622.

- Jackman, Simon. 2000. "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo." *American Journal of Political Science* 44(2):375–404.
- Kim, Jae-On and Charles W. Mueller. 1978. *Factor Analysis: Statistical Methods and Practical Issues*. Sage Publications.
- Laporte, Audrey and Frank Windmeijer. 2005. "Estimation of Panel Data Models with Binary Indicators when Treatment Effects are not Constant Over Time." *Economics Letters* 88:389–396.
- Lewis, David E. 2008. *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance*. Princeton University Press.
- Maskin, Eric and Jean Tirole. 2004. "The Politician and the Judge: Accountability in Government." *American Economic Review* 94(4):1034–1054.
- Mayhew, David. 1991. *Divided We Govern*. Yale.
- Mmuya, Max and Amon Chaligha. 1992. *Towards Multiparty Politics in Tanzania*. Dar es Salaam University Press.
- Mmuya, Max and Amon Chaligha. 1994. *Political Parties and Democracy in Tanzania*. Dar es Salaam University Press.
- Paluck, Elizabeth Levy and Donald P. Green. 2009. "Deference, Dissent, and Dispute Resolution: An Experimental Intervention Using Mass Media to Change Norms and Behavior in Rwanda." *American Political Science Review* 103:622–644.
- Rauch, James E. and Peter B. Evans. 2000. "Bureaucratic Structure and Bureaucratic Performance in Less Developed Countries." *Journal of Public Economics* 75:49–71.
- Tomz, Michael, Jason Wittenberg and Gary King. 2001. "CLARIFY: Software for Interpreting and Presenting Statistical Results. Version 2.0." <http://gking.harvard.edu>. Cambridge, MA: Harvard University.
- Treier, Shawn and Simon Jackman. 2007. "Democracy as a Latent Variable."
- Tsebelis, George. 1989. "The Abuse of Probability in Political Analysis: The Robinson Crusoe Fallacy." *The American Political Science Review* 83(1):77–91.

A Statistical Appendix

A.1 Public Procurement

This portion of the appendix pertains to the audits conducted by Kilimanjaro International in support of the PPRA. For full details of this analysis see Section 3.1.1. Below, we explain in greater detail the construction of the ‘overall compliance’ indicator constructed from the compliance measures reported in Kilimanjaro’s audits and in the PPRA follow-up audits. We additionally present results of an analysis intended to identify the *causal* effects of auditing by isolating trends in patterns of compliance over time.

Index of Compliance Indicators

In Section 3.1.1, we present an analysis of changes in the levels of each of the 13 compliance indicators measured by Kilimanjaro and the PPRA before and after auditing. We additionally present an analysis using a measure of overall compliance. To construct this measure of overall compliance, we rely on a principal components analysis of the 13 different compliance measures.

Principal components analysis (PCA) is a means of collapsing many measures of a single phenomenon (in this case ‘compliance’), into a weighted average of these measures. This weighted average is constructed taking into account the variation in these measures across units. For instance, if all thirteen indicators move in lock step across procuring entities – such that a if procuring entity A has a one standard deviation higher level of compliance than procuring entity B with regards to advertisement, then procuring entity A will also have a one standard deviation higher level of compliance with regards to documentation – then each of these measures will contribute equally to the construction of the weighted average. If, however, 12 of the 13 indicators move in lock step, while the 13th demonstrates a different pattern, the index constructed by PCA will downweight the 13th indicator. The contribution of each indicator to this index is described in Table A.1 in Appendix A.1.

PCA examines how readily the common variation in outcome measures can be explained by multiple orthogonal latent dimensions, or ‘components.’ We construct our overall measure using the first component from the PCA analysis. This component explains 36 percent of the total variation in the data – whereas the second component explains a mere 13 percent of the variation. The overall compliance measure constructed using PCA analysis correlates with the simple average at a level of 0.99 – implying that our results would

be substantively unchanged by using a the PPRA's overall compliance measure, which uses a simple average of compliance scores.

The eigenvectors from the principal components analysis are reported in Table A.1 below. If a variable receives a higher value eigenvector, this implies that it is given greater weight in the index.

Table A.1: Eigenvectors From Principal Components Analysis

Compliance Measure	Eigenvector
Pct. Advertised	0.313
Pct. Comp. Approval	0.355
Tender Board	0.203
Pct. Stnd. Doc.	0.191
Interference	0.331
Contract Imp.	0.195
Pct. Auth. Method	0.308
PMU	0.178
Procurement Plan	0.305
Pct. Publicized	0.259
Quality Oversight	0.279
Pct. Complete Records	0.256
Pct. Compliance Time	0.434

In Section 3.1.1, we regress the outcome measure constructed using PCA on an indicator for whether a given audit is a follow-up audit. If scores are improving following auditing, the coefficient on this indicator should be positive. As is documented above, this coefficient is positive and highly significant. Auditing is associated with a 1.5 standard deviation increase in our overall compliance indicator.

Testing for Time Trends in Compliance

As noted in Section 3.1.1, simple differences in compliance levels over time do not reflect the *causal* effect of auditing. The improvements in compliance scores following auditing that we document above may reflect a general trend towards increasing levels of compliance by procuring entities over time. In the analysis below, we attempt to assess whether or not time trends are confounding our results.

To do so, we regress compare the initial audit scores from the FY 2005/6 audits to the initial audit scores from the FY 2006/7 audits, conducted on *different* procuring entities. If there is a general trend towards increasing levels of compliance – and if these two sets

of procuring entities are sufficiently comparable (see Section 3.1.1 for a discussion) – then one would expect the audits covering FY 2006/7 to reveal significantly higher compliance scores than those covering FY 2005/6.

Figure A.1 reports the difference between the compliance scores reported in the FY 2005/6 and in the FY 2006/7 audits, along with 95 percent confidence intervals. These results were obtained by regressing compliance scores on a variable capturing the year in which the audit took place. In addition to the 13 compliance indicators used by the PPRA, we include the difference in the overall compliance indicator described above.

As with the results discussed above, these regressions incur a multiple comparisons problem. Since we examine differences in a number of indicators, it is to be expected that some differences will be statistically significant. Rather than focusing on differences in a single indicator, one should look for patterns that hold across all or most indicators. Particular attention should be paid to differences in the overall compliance measure, since this measure captures information on the general pattern of differences in compliance scores over time.

Most differences reported in Figure A.1 are small and insignificant. The difference in the overall compliance indicator is estimated to be 0.458 with a 95 percent confidence interval running from -0.712 to 1.63.¹⁰² Exceptions to this general pattern are apparent for the percentage of tenders competed in the regulated amount of time – procuring entities in the second round of audits score, on average, 23.94 percentage points higher on this measure, with a confidence interval from 0.249 to 47.65; for the percentage of tenders with complete records – procuring entities in the second round of audits score, on average, 36.42 percentage points higher on this measure, with a confidence interval from 24.05 to 48.83; and for the creation of a PMU – procuring entities in the second round of audits score, on average, 64.4 points lower on this indicator, with a confidence interval from -43.40 to -85.41.

In general, however, there does not appear to be a clear pattern of improved compliance over time. It seems, therefore, that the increased levels of compliance between the initial round of auditing and the follow-up audits documented above are primarily the result of procuring entity responses to the audit reports. Procuring entities appear to adjust their procurement behavior in response to documented shortcomings in the audit reports. There does not appear to be a general trend of improvements in compliance over time. These

¹⁰²The range of this indicator – from -4.88 to 3.56 – is much smaller than that for the other indicators – which range from 0 to 100. This explains the relatively narrow confidence interval for the difference in overall compliance.

results suggest that TCP activities, other than auditing, did not drive changes in procuring entity behavior during the 2006/7 fiscal year.

The absence of a clear time trend between fiscal year 2005/6 and fiscal year 2006/7 need not imply that non-auditing activities – e.g., the strengthening of the PPRA, stakeholder fora – were ineffective. First, we can only isolate a trend in compliance behavior over the interval covered by the fiscal years 2005/6 and 2006/7. Since these activities took place in 2007 and 2008, it is possible that there was simply insufficient time for an effect to materialize in our estimates. Second, our estimates can only discern a time trend. We cannot control for other changes that might obscure the effect of training activities and stakeholder fora. Third, it is possible that these activities had effects on other areas of procurement that are not captured in the compliance measures used here. However, the absence of a clear improvement in the initial audit scores over time does suggest that training and stakeholder fora did not have a large or immediate effect on procuring entity compliance.

A.2 Rule of Law

This portion of the appendix pertains to the ICITAP's training of the PCCB. For a full discussion of our evaluation of this training, see Section 3.2.1. Below, we present our decomposition of the effects of different forms of training provided by ICITAP. Specifically, we examine the differing effects of Advanced, Intermediate, and Basic training in economic and financial crimes.

Decomposing the Effect of ICITAP Training Sessions

In Section 3.2.1, we examine the average effect of ICITAP training sessions on a variety of outcome measures pertaining to the PCCB's activities in combating corruption. These outcomes include the convictions as a percentage of cases brought to trial, cases prosecuted or referred to the DPP as a percentage of open cases, and new cases opened as a percentage of referrals. These results did not reveal an effect of ICITAP's training. However, it remains possible that this average result masks heterogeneity in the effect of training across different forms of training that were administered.

We test for such heterogeneous effects below. We examine the effectiveness of Advanced, Intermediate, and Basic training in economic and financial crimes. We do not test for effects from training in criminal records management or internal affairs, as these forms of training pertain less clearly to our outcome measures. Our model to test for such effects

is as follows:

$$\begin{aligned}\Delta y_{i,t} = & \beta_0 + \beta_1 \Delta fraction\ advanced_{i,t-a} + \\ & \beta_2 \Delta fraction\ advanced_{i,t-a-1} + \dots + \beta_{b+1} \Delta fraction\ trained_{i,t-a-b} + \\ & \gamma_1 \Delta fraction\ intermediate_{i,t-a} + \gamma_2 \Delta fraction\ intermediate_{i,t-a-1} + \dots + \\ & \gamma_{b+1} \Delta fraction\ intermediate_{i,t-a-b} + \lambda_1 \Delta fraction\ basic_{i,t-a} + \lambda_2 \Delta fraction\ basic_{i,t-a-1} + \\ & \lambda_{b+1} \Delta fraction\ basic_{i,t-a-b} + \epsilon_{i,t}\end{aligned}$$

where i represents region i , t represents month t , a and b specify the number of lags to include in the regression, and Δ is the first-difference operator. $\epsilon \sim N(0, \Sigma)$, where Σ is specified to allow each region's disturbance to be correlated over time, and $E[\epsilon_{i,t} | \Delta X_{i,t-a} \dots \Delta X_{i,t-a-b}] = 0$. y represents one of the outcome measures described above.

The regressors $\Delta fraction\ advanced$, $\Delta fraction\ intermediate$, and $\Delta fraction\ basic$ are defined in a similar manner to $\Delta fraction\ trained$, described in Section 3.2.1. Each capture the fraction of districts in a given region that receive training in a given month. Each variable assumes values between zero and one. And each only takes non-zero values in the month training is provided.

The results of one such regression on the percentage of court cases leading to convictions are presented graphically in Figure A.2. The regressand is defined as the ratio of convictions to total on-going court cases. Coefficient estimates for each type of training are presented in the graphs bearing the corresponding title. Point estimates are presented as diamonds, while 95 percent confidence intervals are represented as lines.

Very little evidence of an effect of any type of training can be seen in Figure A.2. Very few coefficient estimates approach statistical significance at 95 percent level. Nor is the direction of coefficient estimates consistent for any type of training. Based on this evidence, we are unable to reject the null hypothesis that ICITAP training does not affect the ratio of convictions to court cases.

We test for an effect of the different forms of training on the distribution of cases brought for prosecution in Figures A.4 and A.3. As discussed above, superior technical skill on the part of the PCCB may lead prosecutors to bring more difficult cases to trial. If this is the case, we would expect training to affect the percentage of cases prosecuted and/or cases referred to the DPP. We regress these measures against training – where training is lagged between 4 to 8 months.

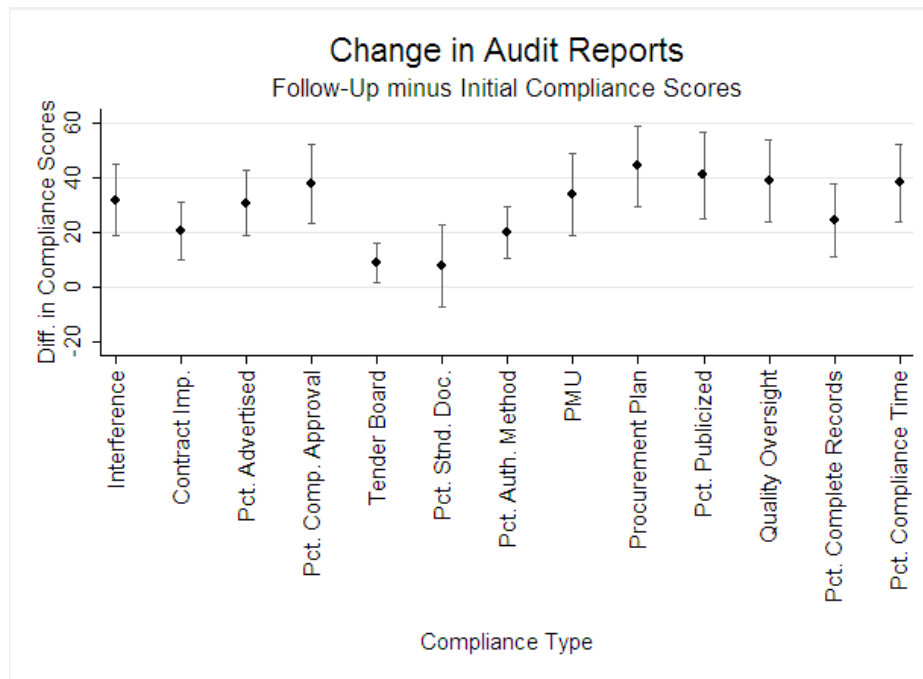
Little evidence of any effect from training is evident in Figures A.4 or A.3. Coefficient

estimates are highly inconsistent – sometimes assuming positive and other time negative values. Very few coefficient estimates approach statistical significance at the 95 percent level.

Finally, we examine the effect of each type of training discussed above on the ratio of new cases to referrals. Improvements in its technical capabilities may make the PCCB more willing to initiate difficult investigations that would previously be dropped. If so, we would expect training to increase this ratio. Coefficient estimates for this regression are presented in Figure A.5.

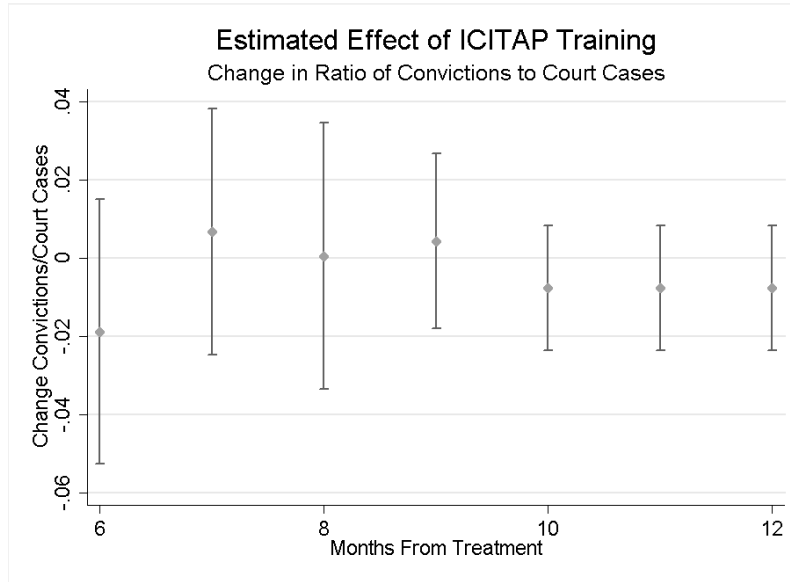
Again, few patterns are evident in the coefficient estimates. Coefficient estimates on advanced economic and financial crimes training generally appear to be positive – indicating an increasing ratio of new cases to referrals. But, these estimates are never significant. Estimates for the other forms of training are insignificant and vary in sign. We cannot conclude, based on this evidence, that ICITAP training improved PCCB performance.

Figure 1: Average Change in Compliance Scores



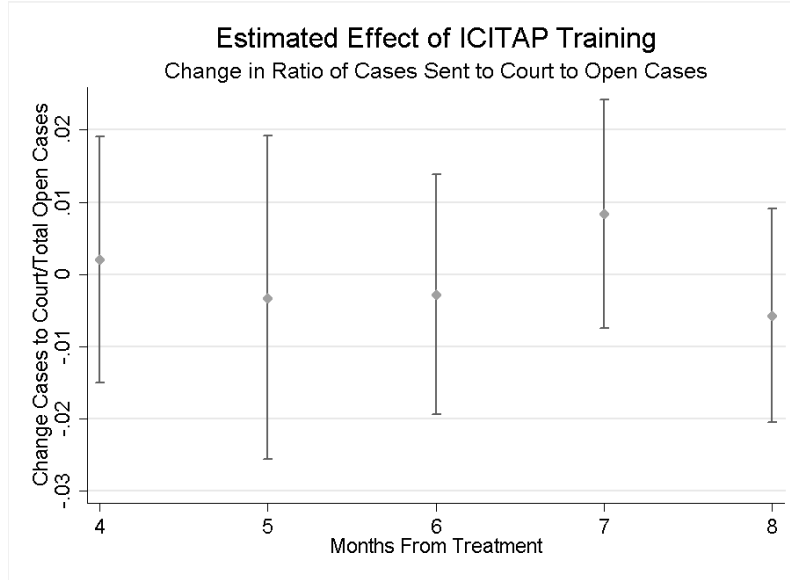
Average differences between follow-up and initial audit compliance scores, across the 13 compliance measures recorded by the PPRA. Positive values denote an improvement in performance as measured by these indicators between the initial and follow-up audits. Negative numbers indicate the reverse. Averages are taken across 19 Fiscal Year 2005/6 initial and 2006/7 follow-up audits and 18 Fiscal Year 2006/2007 initial and 2007/2008 follow-up audits. Dots denote the average difference, while whiskers denote 95 percent confidence intervals. If the whiskers cross the zero line, we cannot say that these results are statistically significant at the 95 percent level.

Figure 2: Effect of Training on Conviction Rate



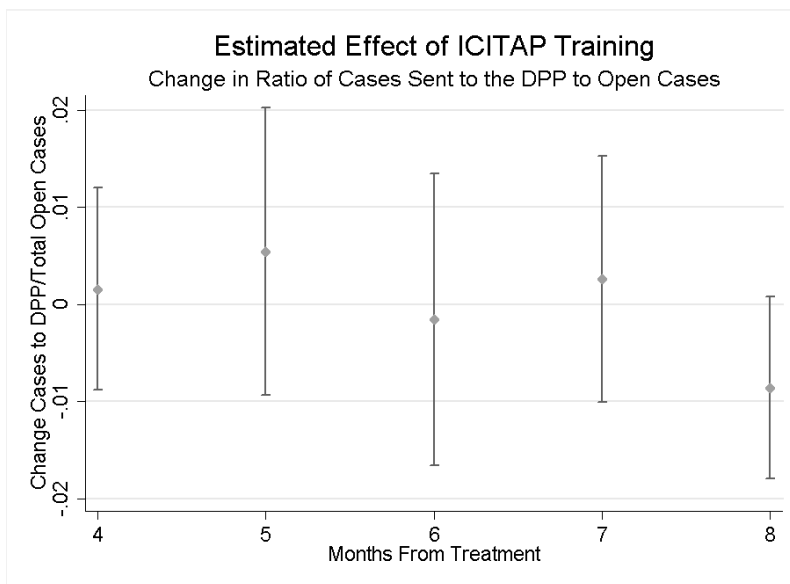
Coefficient estimates from a regression of the change in the ratio of convictions to on-going court cases. The regressand is the number of convictions divided by the number of on-going court cases in a given region in a given month, as reported by the PCCB. The regressors are measures of the fraction of districts within a given region that receive training from the ICITAP advanced and basic economic and financial crimes courses, the intermediate fraud and financial crimes courses, the internal affairs development courses, or the supervision and management courses in a given month. We include lags of the regressor of between 6-12 months in our specification. Coefficient values represent the estimated difference in the rate of change in the ratio of convictions to on-going court cases in regions in which all districts were trained relative to regions in which no districts were trained. Coefficient estimates are represented as diamonds in the above. The 95 percent confidence intervals are presented as lines. If the 95 percent confidence interval crosses zero, we cannot say that training as a statistically significant effect on the ratio of convictions to on-going court cases at the 95 percent level. All standard errors are clustered by region.

Figure 3: Effect of Training on Trial Rate



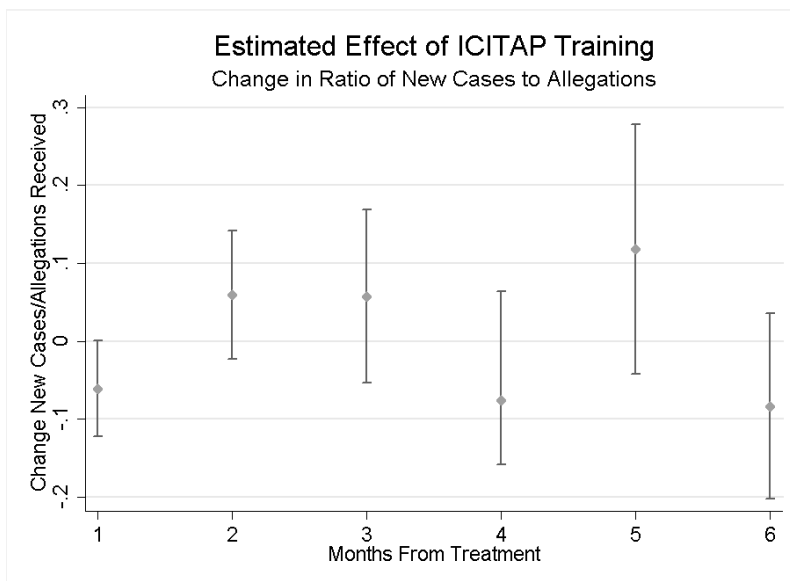
Coefficient estimates from a regression of the change in the ratio of cases prosecuted (either by the DPP or the PCCB) to open cases (new cases + on-going cases). The regressand is the number of cases prosecuted divided by the number of open cases in a given region in a given month, as reported by the PCCB. The regressors are measures of the fraction of districts within a given region that receive training from the ICITAP advanced and basic economic and financial crimes courses, the intermediate fraud and financial crimes courses, the internal affairs development courses, or the supervision and management courses in a given month. We include lags of the regressor of between 4-8 months in our specification. Coefficient values represent the estimated difference in the rate of change in the ratio of prosecutions to open cases in regions in which all districts were trained relative to regions in which no districts were trained. Coefficient estimates are represented as diamonds in the above. The 95 percent confidence intervals are presented as lines. If the 95 percent confidence interval crosses zero, we cannot say that training has a statistically significant effect on the ratio of prosecutions to open cases at the 95 percent level. All standard errors are clustered by region.

Figure 4: Effect of Training on DPP Referrals



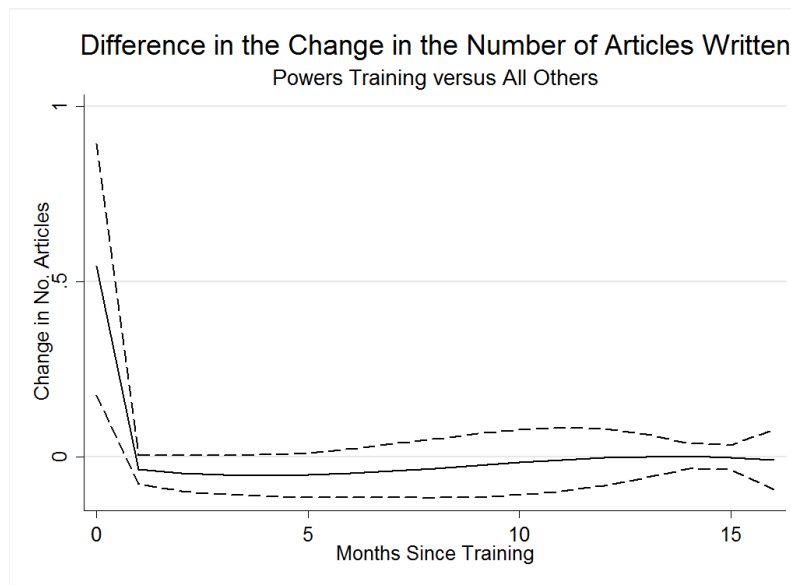
Coefficient estimates from a regression of the change in the ratio of cases referred to the DPP to open cases (new cases + on-going cases). The regressand is the number of cases referred to the DPP divided by the number of open cases in a given region in a given month, as reported by the PCCB. The regressors are measures of the fraction of districts within a given region that receive training from the ICITAP advanced and basic economic and financial crimes courses, the intermediate fraud and financial crimes courses, the internal affairs development courses, or the supervision and management courses in a given month. We include lags of the regressor of between 4-8 months in our specification. Coefficient values represent the estimated difference in the rate of change in the ratio of DPP referrals to open cases in regions in which all districts were trained relative to regions in which no districts were trained. Coefficient estimates are represented as diamonds in the above. The 95 percent confidence intervals are presented as lines. If the 95 percent confidence interval crosses zero, we cannot say that training has a statistically significant effect on the ratio of DPP referrals to open cases at the 95 percent level. All standard errors are clustered by region.

Figure 5: Effect of Training on Case Uptake



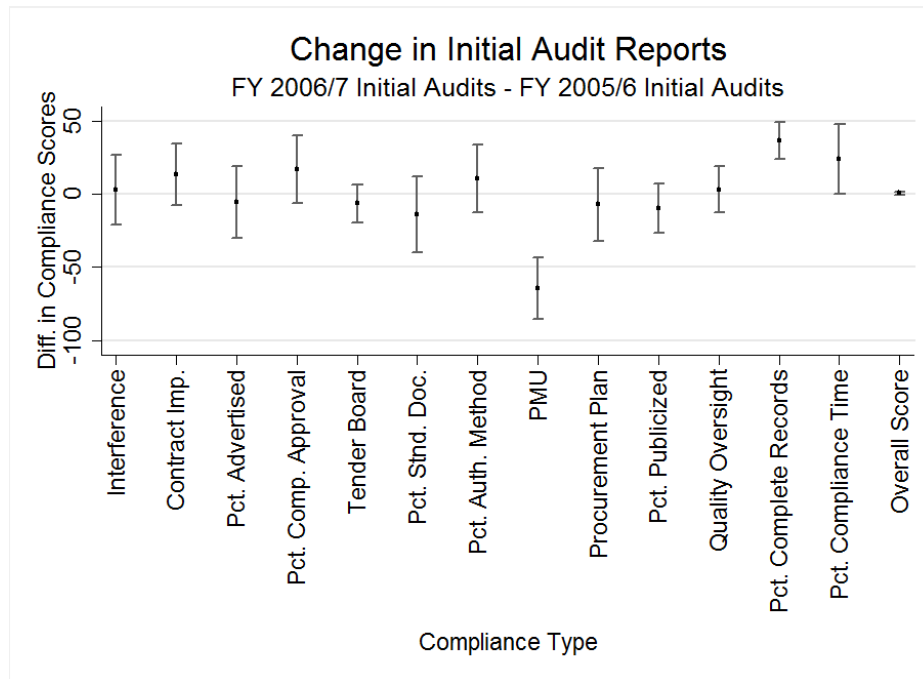
Coefficient estimates from a regression of the change in the ratio of new cases to allegations on training. The regressand is the number of new cases opened divided by the number of allegations received in a given region in a given month, as reported by the PCCB. The regressors are measures of the fraction of districts within a given region that receive training from the ICITAP advanced and basic economic and financial crimes courses, the intermediate fraud and financial crimes courses, the internal affairs development courses, or the supervision and management courses in a given month. We include lags of the regressor of between 1-6 months in our specification. Coefficient values represent the estimated difference in the rate of change in the ratio of new cases to referrals in regions in which all districts were trained relative to regions in which no districts were trained. Coefficient estimates are represented as diamonds in the above. The 95 percent confidence intervals are presented as lines. If the 95 percent confidence interval crosses zero, we cannot say that training as a statistically significant effect on the ratio of new cases to referrals at the 95 percent level. All standard errors are clustered by region.

Figure 6: Simulated Difference btwn Powers-trained and Others



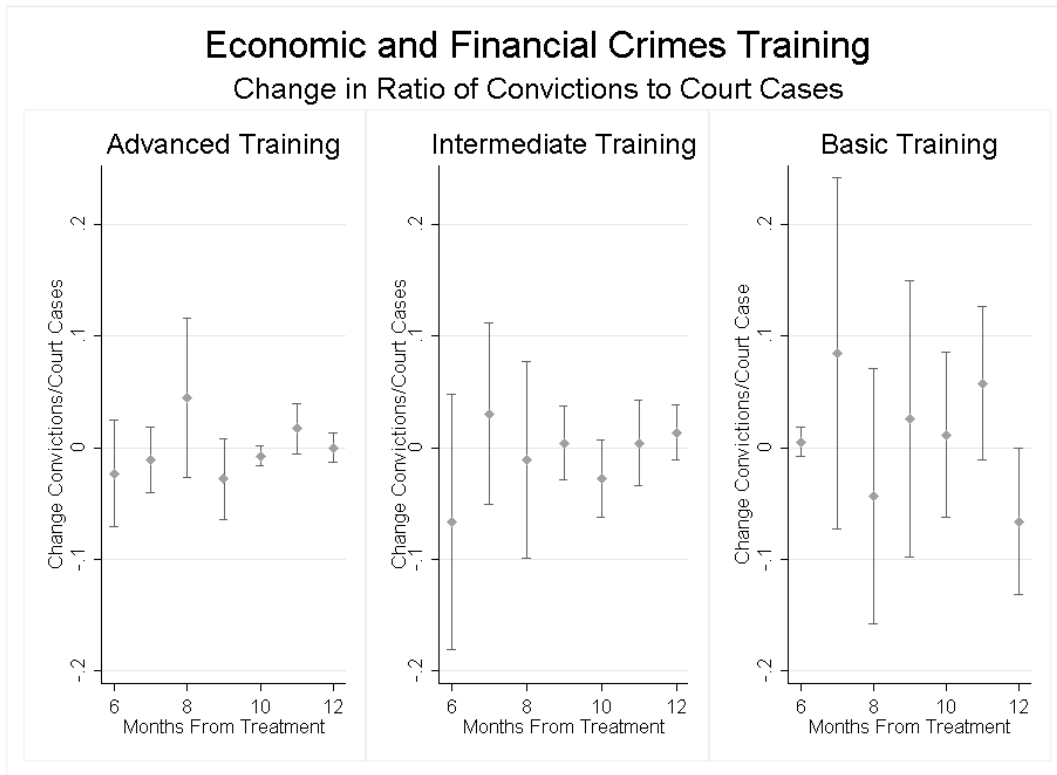
Simulations from model 2 in Table 8. The solid line represents the predicted difference between the change in the number of articles produced by a journalist who received the powers training in time $t = 1$ and the change in the number of articles produced by all other journalists. Dashed lines represent 95 percent confidence intervals. Results indicate an increase in the rate of change of the number of articles produced by Powers trained journalists – relative to all other journalists – in the month of training. Following training, the rate of change is statistically identical in the two groups. Thus, training produces a permanent positive shift in the number of articles produced by trained journalists. Results are produced using the CLARIFY (Tomz, Wittenberg and King, 2001) software run from Stata 11.

Figure A.1: Change in Compliance Scores Not Attributable to Auditing



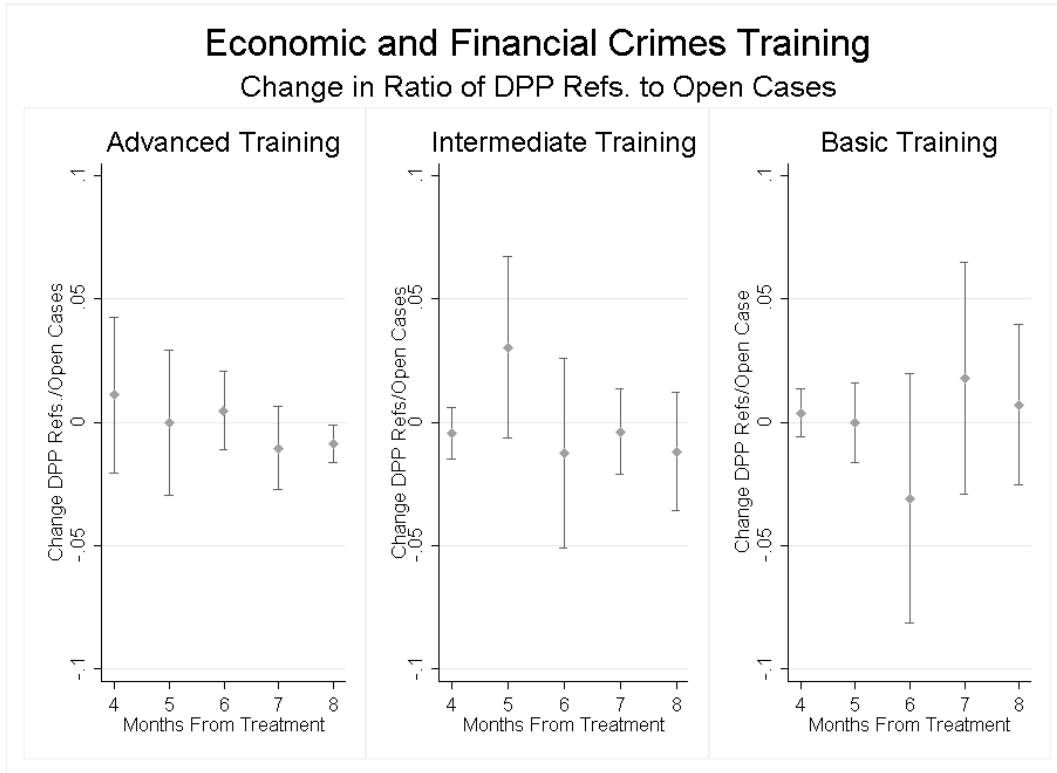
Average differences between initial fiscal year 2006/7 compliance scores and initial fiscal year 2005/6 compliance scores, across the 13 compliance measures recorded by the PPRA and the overall compliance measure constructed above. Positive values denote an improvement in performance over time, as measured by these indicators. Negative numbers indicate the reverse. Averages are taken across 19 Fiscal Year 2005/6 and 18 2006/7 initial audits. Dots denote the average difference, while whiskers denote 95 percent confidence intervals. If the whiskers cross the zero line, we cannot say that these results are statistically significant at the 95 percent level.

Figure A.2: Disaggregated Effect of Training on Conviction Rate



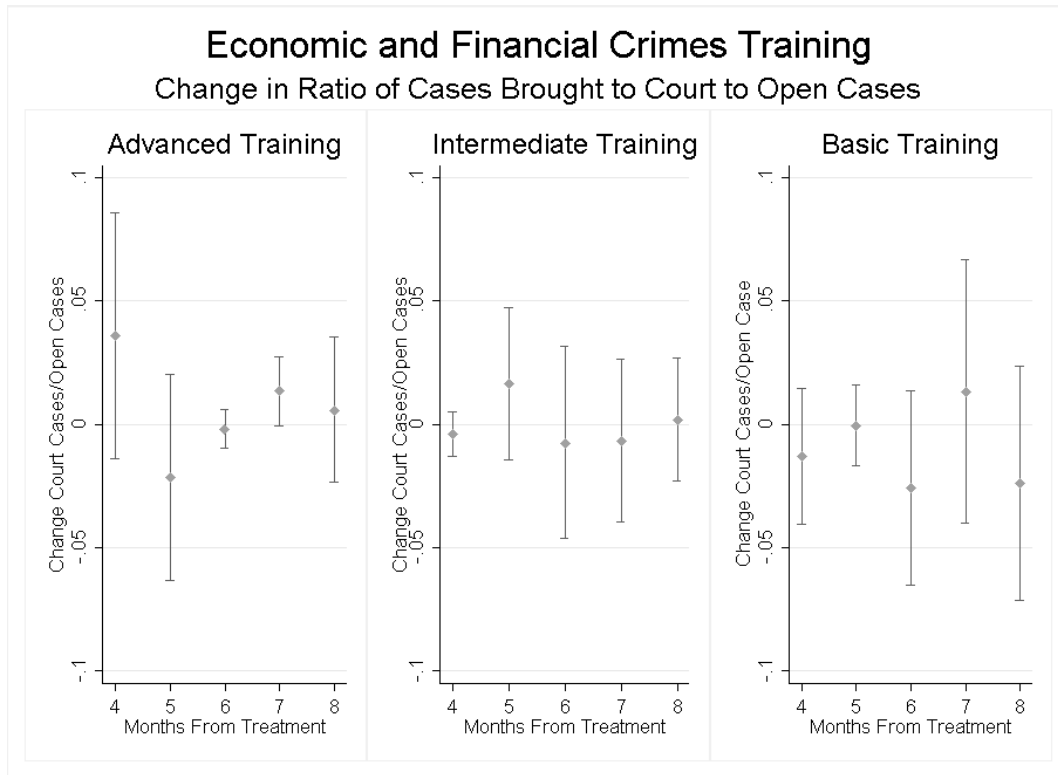
Coefficient estimates from a regression of the change in the ratio of convictions to on-going court cases. The regressand is the number of convictions divided by the number of on-going court cases in a given region in a given month, as reported by the PCCB. The regressors are measures of the fraction of districts within a given region that receive training from the ICITAP advanced and basic economic and financial crimes courses, and the intermediate fraud and financial crimes courses in a given month. Coefficient values are allowed to vary by the type of training received. We include lags of the regressor of between 6-12 months in our specification. Coefficient values represent the estimated difference in the rate of change in the ratio of convictions to on-going court cases in regions in which all districts were trained relative to regions in which no districts were trained. Coefficient estimates are represented as diamonds in the above. The 95 percent confidence intervals are presented as lines. If the 95 percent confidence interval crosses zero, we cannot say that training as a statistically significant effect on the ratio of convictions to on-going court cases at the 95 percent level. All standard errors are clustered by region.

Figure A.3: Disaggregated Effect of Training on DPP Referrals



Coefficient estimates from a regression of the change in the ratio of cases referred to the DPP to open cases (new cases + on-going cases). The regressand is the number of cases referred to the DPP divided by the number of open cases in a given region in a given month, as reported by the PCCB. The regressors are measures of the fraction of districts within a given region that receive training from the ICITAP advanced and basic economic and financial crimes courses, and the intermediate fraud and financial crimes courses in a given month. Coefficient estimates are allowed to vary by the type of training received. We include lags of the regressor of between 4-8 months in our specification. Coefficient values represent the estimated difference in the rate of change in the ratio of DPP referrals to open cases in regions in which all districts were trained relative to regions in which no districts were trained. Coefficient estimates are represented as diamonds in the above. The 95 percent confidence intervals are presented as lines. If the 95 percent confidence interval crosses zero, we cannot say that training as a statistically significant effect on the ratio of DPP referrals to open cases at the 95 percent level. All standard errors are clustered by region.

Figure A.4: Disaggregated Effect of Training on Trial Rate



Coefficient estimates from a regression of the change in the ratio of cases prosecuted (either by the DPP or the PCCB) to open cases (new cases + on-going cases). The regressand is the number of cases prosecuted divided by the number of open cases in a given region in a given month, as reported by the PCCB. The regressors are measures of the fraction of districts within a given region that receive training from the ICITAP advanced and basic economic and financial crimes courses, and the intermediate fraud and financial crimes courses in a given month. Coefficient estimates are allowed to vary by the type of training received. We include lags of the regressor of between 4-8 months in our specification. Coefficient values represent the estimated difference in the rate of change in the ratio of prosecutions to open cases in regions in which all districts were trained relative to regions in which no districts were trained. Coefficient estimates are represented as diamonds in the above. The 95 percent confidence intervals are presented as lines. If the 95 percent confidence interval crosses zero, we cannot say that training as a statistically significant effect on the ratio of prosecutions to open cases at the 95 percent level. All standard errors are clustered by region.

Figure A.5: Disaggregated Effect of Training on Case Uptake



Coefficient estimates from a regression of the change in the ratio of new cases to allegations on training. The regressand is the number of new cases opened divided by the number of allegations received in a given region in a given month, as reported by the PCCB. The regressors are measures of the fraction of districts within a given region that receive training from the ICITAP advanced and basic economic and financial crimes courses, and the intermediate fraud and financial crimes courses in a given month. Coefficient values are allowed to vary depending on the type of training received. We include lags of the regressor of between 1-4 months in our specification. Coefficient values represent the estimated difference in the rate of change in the ratio of new cases to referrals in regions in which all districts were trained relative to regions in which no districts were trained. Coefficient estimates are represented as diamonds in the above. The 95 percent confidence intervals are presented as lines. If the 95 percent confidence interval crosses zero, we cannot say that training as a statistically significant effect on the ratio of new cases to referrals at the 95 percent level. All standard errors are clustered by region.

A.3 Journalist Training

This portion of the appendix pertains to our evaluation of PACT-Tanzania’s training of investigative journalists. For further details on our evaluation, see Section 3.4.1. In this appendix, we describe the data used, describe our method for constructing the measure of article quality used in the evaluation, and present further results from our analysis not contained in the main body of the text.

Data Description: The Quantity of Articles Produced

Recall that we rely on a collection of press clippings from Dar es Salaam circulated newspapers maintained by PACT as a source for our data. This dataset contains all articles relating to good governance and corruption released in Dar es Salaam-circulated newspapers between January 2007 and November 2008 – excepting the months of April and May 2008. Table A.2 lists the set of papers contained in our dataset.

We match all articles with a byline according to the authors name, so that we are left with a panel of each author’s output over the period. More precisely, we examine the number of articles produced by a given author in a given month. Figure A.6 plots a histogram of the distribution of the number of articles produced per month. As is readily evident from this figure, most authors produce a very small number of articles and the modal observation is zero.

Our analysis in Section 3.4.1 does not examine the number of articles produced per journalist per month, but rather focuses on the rate of change in article production. By focusing on this rate of change, we are able to control for journalist specific factors that may affect both selection into training and the quantity of journalistic output. Figure A.7 plots a histogram of the distribution of the rate of change in the number of articles produced per journalist per month. Note that, again, this distribution is closely centered on zero.

Data Description: The Quality of Articles Produced

In this portion of the appendix, we discuss the statistical methods used to construct the article quality index used to evaluate the effect of training on the quality of articles journalists produce. The details of this analysis are related in Section 3.4.1.

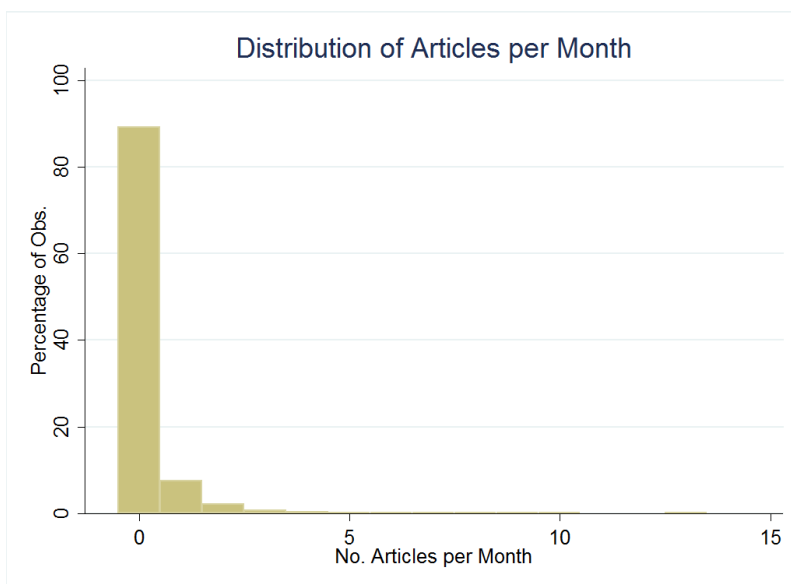
Recall that we construct this index for a random sample of 55 authors drawn from the set of all authors included in our dataset. We had four evaluators from EcomResearch read each of the articles produced by these authors and independently code their responses to a number of questions pertaining to article quality. In total, 374 articles were thus evaluated.

Table A.2: List of Newspapers Covered in the Data

Alasiri
An - Nuur
Business Standard
Business Times
Changamoto
Daily News
Dar Leo
East African
Habari Leo
Hoja
Ijumaa
Kulikoni
Maisha wikiendi
Majira
Msemakweli
Mtanzania
Mwana halisi
Mwanachi
Mzalendo
Nipashe
Nyakati
Rai
Raia Mwema
Sauti ya watu Tanzania
Sunday Citizen
Sunday News
Sunday Observer
Tanzania Daima
The African
The Citizen
The East African
The Express
The Guardian
This Day
Uhuru
Uwazi

We construct an index of article quality making use of the information provided in enumerator responses to these questions. More specifically, we estimate article quality making use of an item response model. Item response models are widely used in psychometrics

Figure A.6: Distribution of Articles per Journalist per Month



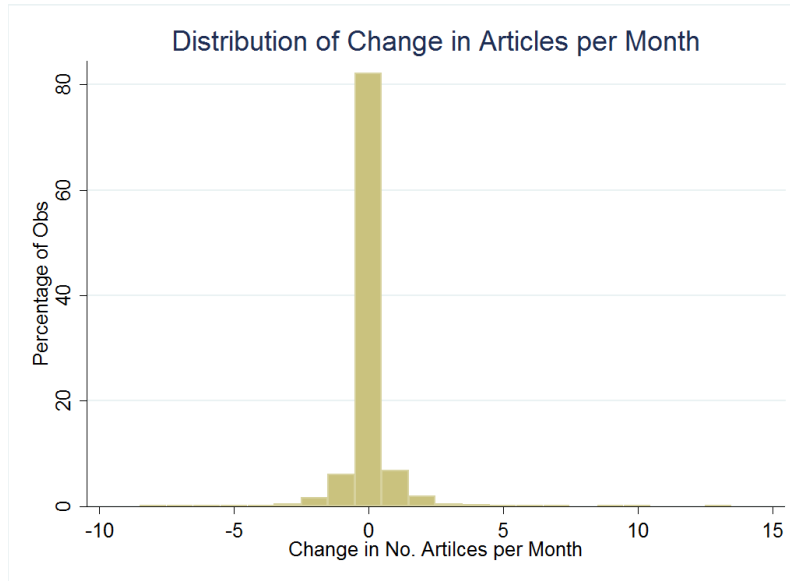
The distribution of the number of articles published by a given journalist in a given month in the period from January 2007 through November 2008 (excluding April and May 2008). The percentage of all observations is given on the y-axis. The number of articles per month is given on the x-axis.

and political science to scale individuals or units along an unobserved (latent) dimension. For instance, in psychometrics, these models are used to assess student capabilities based on responses to standardized tests. In political science, these models are used to assess the ideological leanings of legislators based on roll call voting outcomes.¹⁰³

In an item response model, observed outcomes are treated as a function of the latent dimension. So the probability that student i gets question j correct on a test $y_{i,j} = 1$, can be modeled as $Pr(y_{i,j} = 1) = \text{logit}^{-1}(\beta_{0j} + \beta_{1j} \text{ability}_i + \epsilon_{i,j})$. Here, β_{0j} measures the difficulty of question j – it affects the average number of students who get the question correct. β_{1j} is known as the discrimination parameter – it measures the degree to which students with higher ability are more likely to get the question correct. With a sufficient number of questions and of students, it is possible to estimate β_{0j} , β_{1j} and ability_i based

¹⁰³For political science applications, see Clinton, Jackman and Rivers (2004) on identification of legislator ideologies from roll call votes, see Treier and Jackman (2007) on constructing an index of democracy, see Clinton and Lewis (2008) on the use of item response theory to extract a measure of agency ideology based on expert surveys.

Figure A.7: Distribution of the Change in the No. of Articles per Journalist per Month



The distribution of the change in the number of articles published by a given journalist in a given month during the period from January 2007 through November 2008 (excluding April, May and June 2008). The percentage of all observations is given on the y-axis. The change in the number of articles is given on the x-axis.

on patterns of variation in the data (Gelman and Hill, 2006).

The resultant index (in the above example, *ability*) has a substantive interpretation – an increase in the value of 1 in the value of the index increases the probability a given outcome is realized (the student answers a question correctly) by a factor measured by the discrimination parameter (β_{1j}). This consistent across measures, such that moving from an index score of 0 to a score of 1 has the same effect as moving from 1 to 2. However, the index is only defined up to an affine transformation – the index may be multiplied by or added to a constant without affecting the interpretation. Any such transformation would also shift the value of the other parameter (i.e., β_{0j} and β_{1j}) to offset this change (Gelman and Hill, 2006).

We construct an item response model using the following series of equations:¹⁰⁴

¹⁰⁴Before estimating these equations, we standardize the ‘overall quality’ measure for each enumerator by subtracting his (each enumerator is male) mean score from each entry and dividing by the standard deviation of his scores. Thus, each enumerator’s overall score is of mean zero and standard deviation one and reflects a given article’s quality *relative to the average quality in the sample*. This step is necessary because most of

$$\begin{aligned}
overall_{i,j} &= \kappa_{0j} + \kappa_{1j} QualityIndex_i + \epsilon_{i,j} \\
clarity_{i,j} &= \text{logit}^{-1}(\beta_{0j} + \beta_{1j} QualityIndex_i + \beta_{2j} swahili_{i,j} + \beta_{3j} technical_{i,j} + \mu_{i,j}) \\
context_{i,j} &= \text{logit}^{-1}(\gamma_{0j} + \gamma_{1j} QualityIndex_i + \gamma_{2j} technical_{i,j} + \eta_{i,j}) \\
sources_{i,j} &= \text{logit}^{-1}(\lambda_{0j} + \lambda_{1j} QualityIndex_i + \lambda_{2j} technical_{i,j} + \omega_{i,j}) \\
grammar_mistakes_{i,j} &= \text{poisson}^{-1}(\zeta_{0j} + \zeta_{1j} QualityIndex_i + \zeta_{2j} swahili_{i,j} + u_{i,j}) \\
num_sources_{i,j} &= \text{poisson}^{-1}(\psi_{0j} + \psi_{1j} QualityIndex_i + \psi_{2j} technical_{i,j} + e_{i,j})
\end{aligned}$$

where i denotes article i and j denotes enumerator j . The coefficients $\kappa_{1j} - \psi_{1j}$ are discrimination parameters. Thus, each article is assigned a score on our quality index and each reviewer may be more or less responsive to the underlying quality dimension. We are also able to control for exogenous factors that may effect coding decisions – the language the article is written in and whether or not the article contains technical content. And $\epsilon_{i,j}$, $\mu_{i,j}$, $\eta_{i,j}$, $\omega_{i,j}$, $u_{i,j}$, and $e_{i,j}$ are all disturbance terms.

We estimate this system of equations using a Markov Chain Monte Carlo (MCMC) estimation technique run from R 2.11.0 and WinBUGS 1.4.¹⁰⁵ We set diffuse priors on each of the coefficients in the models (mean zero variance 100), to ensure our priors do not influence the final results. The values of the quality index are drawn from a prior distribution with mean zero and variance one. As is common practice with such models, we set two articles as anchoring points to ensure that the index is coded such that higher values are

the enumerators tended to cluster their scores in the middle categories. Three of the four enumerators did not score any article a 5 (among the top 20% of articles they had ever read), and several enumerators gave only few scores of 1 (among the bottom 20% of articles they had ever read). Because of this issue, it was not possible to estimate ordered models in the system of equations below. Instead we normalize the overall quality measure and estimate a linear model.

¹⁰⁵MCMC estimation is a technique wherein parameter values are randomly sampled from a prior distribution, the prior distribution is corrected to allow for better approximations of true parameter values, and a new set of parameters are randomly sampled from the new distributions. This process occurs iteratively and, as the number of iterations grows large (converges to infinity), the range of parameters sampled can be shown to converge to the uncertainty inherent in the estimation (Gelman and Hill, 2006; Jackman, 2000). (Posterior distributions from MCMC estimation are analogous to confidence intervals from classical techniques.) To establish that the MCMC algorithm has converged, it is common practice to estimate parallel series of simulations, and to examine ratio of the variations across simulations to that within each simulation. In general, this ratio should not exceed 1.1 for convergence to be established (Gelman and Hill, 2006). We run 3 parallel simulations for 10,000 iterations each, dropping the first 2,000 iterations from our final estimates to ensure that initial values do not unduely influence our results. The ratio of the variance across simulations to that within simulations is approximately 1 for all parameter values.

associated with higher quality and lower scores with lower quality.¹⁰⁶

Figure A.8 contains smoothed kernel density plots of the discrimination parameters for each enumerator as estimated by this system of equations. Recall that these parameters reflect the responsiveness of each enumerator to the latent quality term. Not surprisingly, the responses to the overall quality question are highly responsive to the quality index – all enumerators are likely to rate the ‘overall quality’ of an article that scores high on the index above that of an article that scores low on the index. However, this responsiveness varies across enumerators. Enumerators 2 and 4 are highly responsive; while enumerator 3 is substantially less so. Enumerators 1 and 3 are likely to indicate articles that score more highly on the index cite a greater number of sources; while enumerators 2 and 4 are unresponsive to this question. Three out of four enumerators also are likely to rank highly scored articles as better in setting a given story in context.

Somewhat surprisingly, 3 of 4 enumerators are likely to say that higher quality articles are less likely to be understood by a member of the general public. We interpret this to indicate that these articles are written at a higher grade level than less high quality articles. Equally surprisingly, we find that higher ranking articles are likely to contain more grammatical errors than less high ranking ones. This may reflect the greater linguistic complexity of higher ranking articles. Or it may indicate that the enumerators tended to read better articles more closely, and were thus better able to detect any errors in grammar.

Figure A.9 contains a plot of the estimated quality index score for each article, and a 95 percent confidence interval around that score. Unsurprisingly, most index scores are clustered around the midpoint of the index (value 0). Scores vary substantially around this midpoint; however, and the values of high and low scoring articles differ significantly from mid-range scores. The mean quality value is -0.03 with a standard deviation of 0.81. The index ranges from a minimum value of -2.21 to a maximum of 1.79. It is also possible to see that quality scores tend to cluster by article number in Figure A.9. This is because authors are assigned adjacent article numbers, implying that the quality of a given article depends, in part, on the author who drafted it.

¹⁰⁶Recall that the index can be multiplied by -1 and still maintain its meaning, so long as each discrimination parameter is also multiplied by -1.

Figure A.8: Enumerator Responsiveness to Article Quality

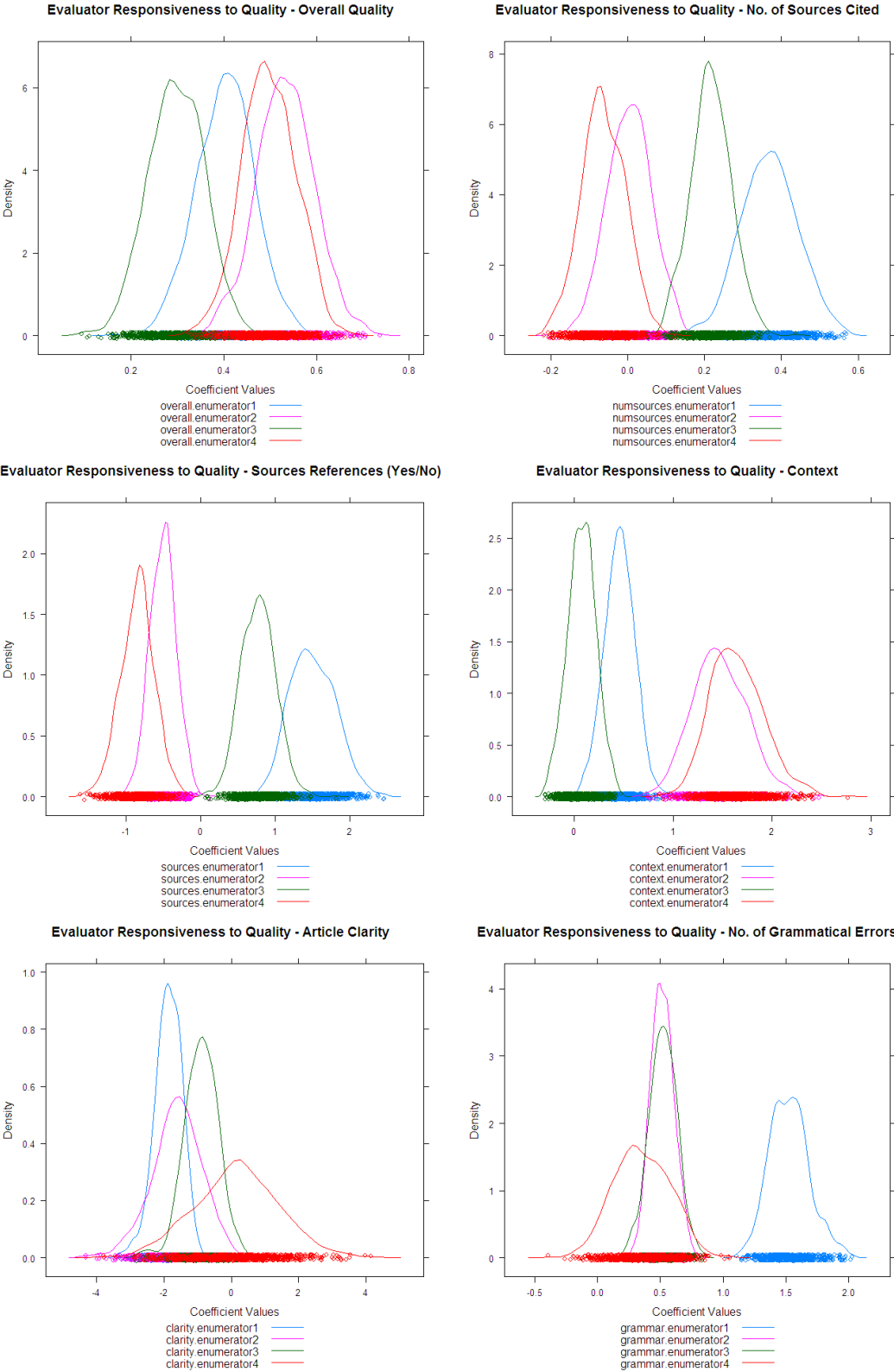
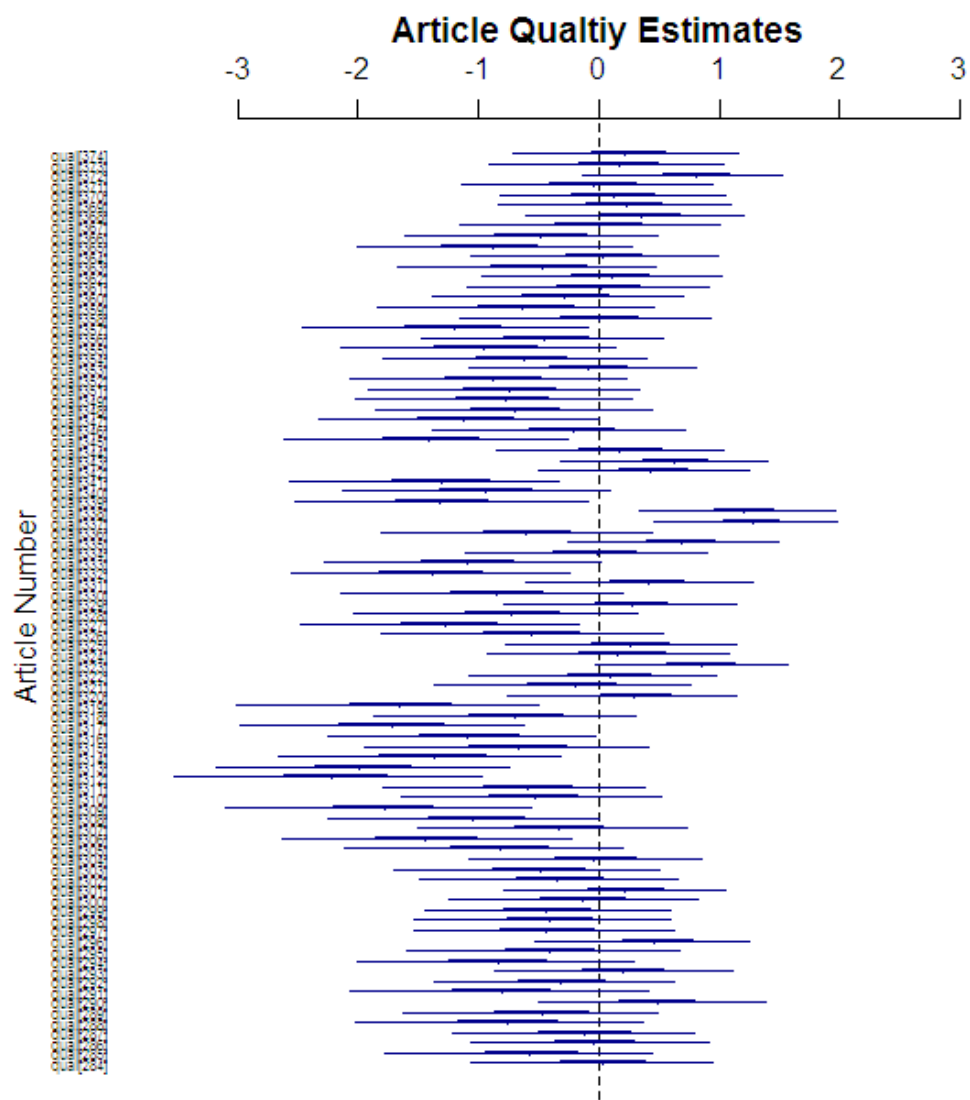
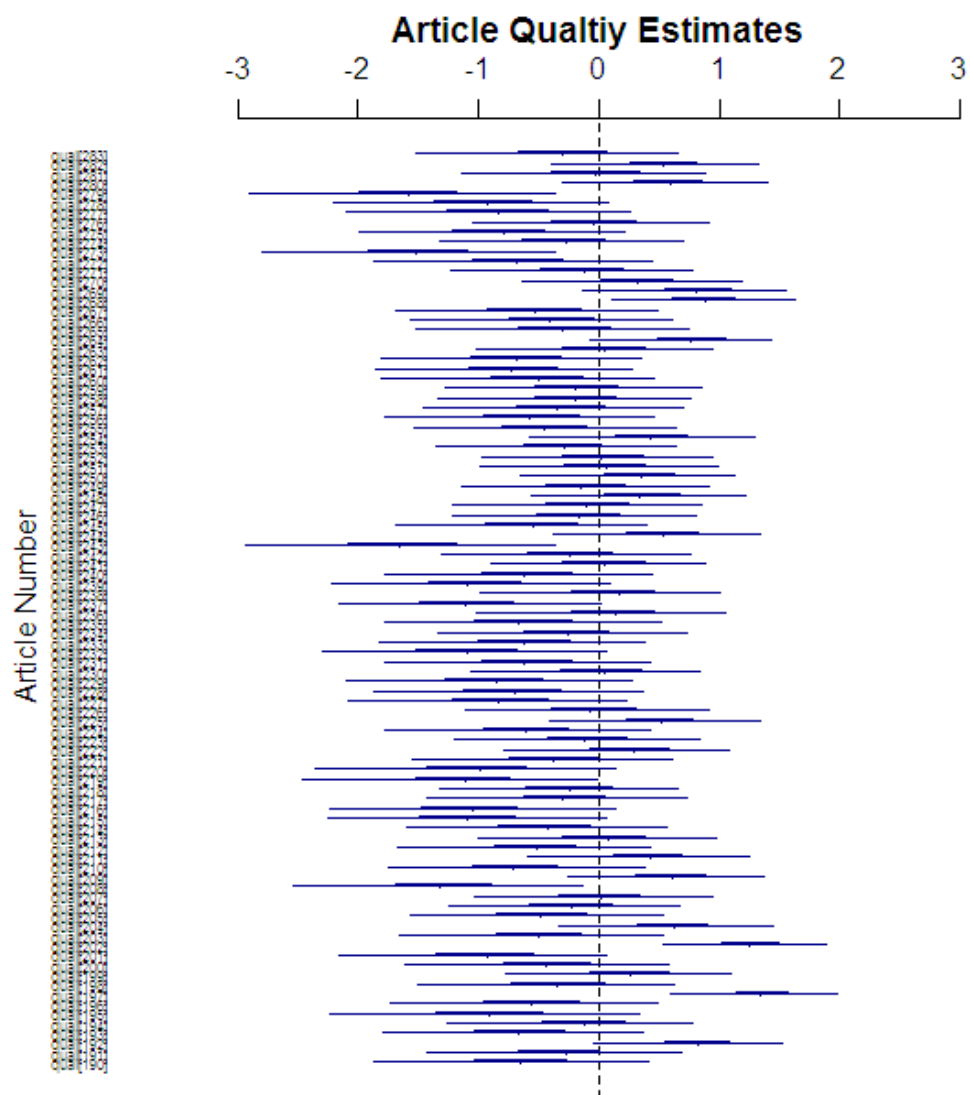
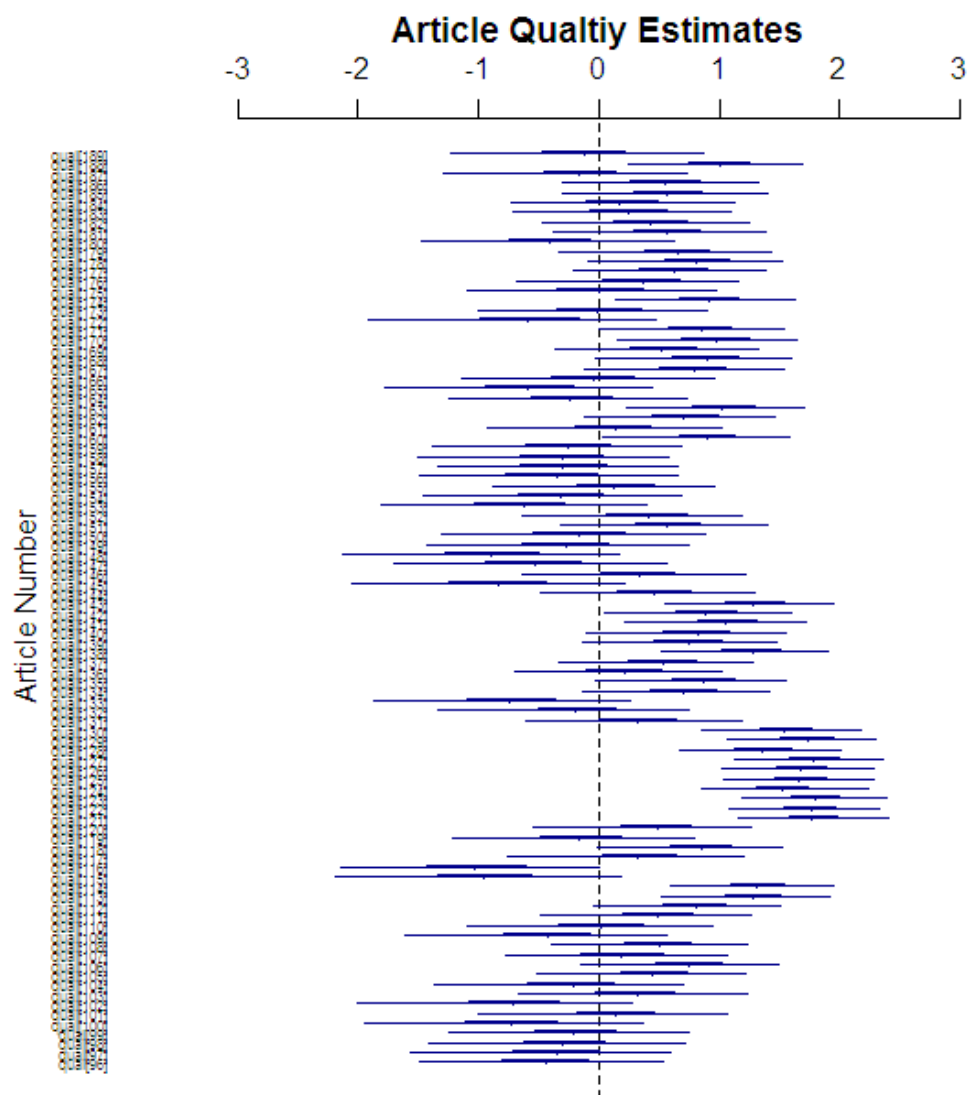
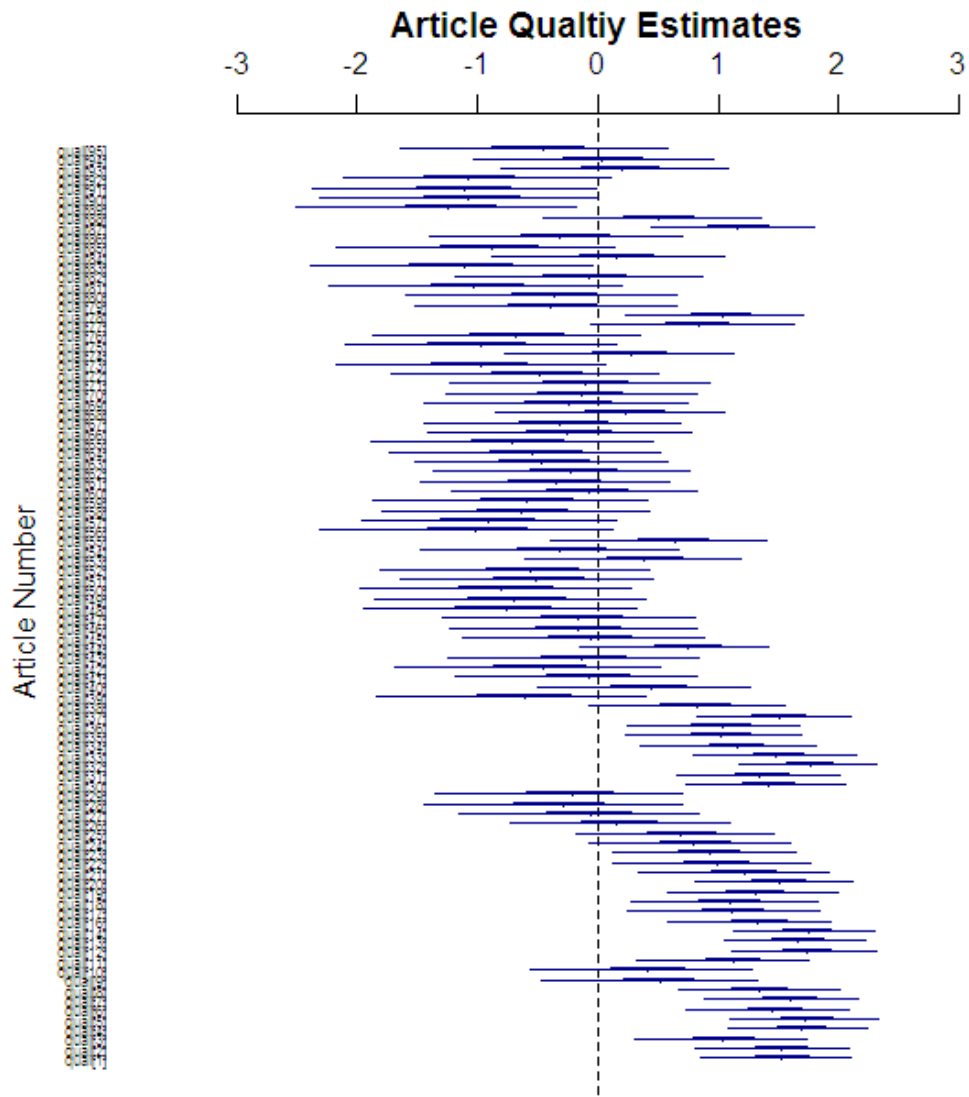


Figure A.9: Estimates of and 95 Percent Confidence Intervals Around Article Quality









Plots of quality index estimates and 95 percent confidence intervals. Quality index scores are denoted on the horizontal axis. Average quality index values across 8,000 simulations are denoted by blue dots. Thick blue lines denote 1 standard deviation around each mean quality index estimate. Thinner blue lines denote 2 standard deviations (95 percent confidence intervals) around these estimates. Each article i is denoted by $qual[i]$ along the left hand vertical axis.

Outcomes: Quantity of Articles Produced

In Section 3.4.1, we examine the effect of training on the quantity of articles produced by a given journalist using a difference-in-differences analysis. Recall that, in that Section, we find that the training session conducted by Jonathan Powers had a large and robust effect on the number of articles produced by trained journalists. We also found that the Lake Zone training session showed some evidence of a positive effect. But, noted that these findings are not robust.

Table A.3 presents the results of our robustness checks focusing on the Lake Zone training session. The outcome measure in all these regressions is the change in the number of articles produced by a given journalist in a given month. The Δ training regressor is a simple binary indicator that takes the value 1 in the month a journalist is trained. The Δ Lake Zone Batch and Δ Powers regressors are similarly defined. In Column 1, we compare those trained in the Lake Zone session to all other trained journalists. In Column 2, we compare those trained in the Lake Zone batch to all other journalists not trained in the Powers session. Finally, in Column 3, we compare those trained in the Lake Zone batch to all other journalists, except those trained in the Powers session. This last specification also examines the changing effects of training over time. In no specification does the Lake Zone training reveal a significant effect on journalist output.

Outcomes: Quality of Articles Produced

In Section 3.4.1, we analyze the effect of PACT's training on the quality of articles produced by journalists, using the quality index discussed above. We additionally test whether the effect of training on article quality differs depending on whether or not an article appears in a partisan newspaper. We code newspapers as being either pro-government, pro-opposition or neutral. To construct our codings, we rely on information made available to us by EcomResearch. A complete listing of newspapers appearing in our quality sample and their partisan affiliations is contained in Table A.4.

In Section 3.4.1, we rely on a difference-in-differences specification to assess the effect of training on article quality. But, such a specification prevents us from assessing the effect of training on authors who participate in early training sessions, as these authors are treated as being trained for the duration of our panel. Since a large number of journalists receive training in the early months of 2007, our results may suffer from this omission. However, we are reluctant to rely on alternative specifications because of the danger that sample selection may bias our results.

Table A.3: Effect of Lake Zone Session

	Δ No. Articles		
Δ training	0.065 [-0.077,0.208]	-0.019 [-0.168,0.129]	
Δ Lake Zone Batch	0.144 [-0.125,0.414]	0.230 [-0.043,0.503]	0.260 [-0.172,0.693]
Δ Powers		0.571** [0.188,0.954]	0.551** [0.198,0.905]
t	-0.007 [-0.018,0.003]	-0.008 [-0.019,0.002]	-0.008 [-0.018,0.003]
t ²	0.002*** [0.001,0.003]	0.002*** [0.001,0.003]	0.002*** [0.001,0.003]
t ³	-0.000*** [-0.000,-0.000]	-0.000*** [-0.000,-0.000]	-0.000*** [-0.000,-0.000]
Months since Lake Zone			0.012 [-0.007,0.032]
Months since Lake Zone ²			-0.000 [-0.005,0.005]
Months since Lake Zone ³			-0.000 [-0.001,0.001]
$\hat{\sigma}^2$	0.717	0.717	0.717

Results of a difference-in-differences regression of the change in the number of articles produced by a given journalist in a given month against the change in training status. The first model compares the effect of the Lake Zone batch training to all other forms of training, the second model compares Lake Zone training to all other forms of training except the Powers session, and the third compares the Lake Zone batch to all other journalists (trained and untrained). A cubic polynomial of time controls for trend effects. In the third model, a cubic polynomial of the number of months since training controls for lagged effects. 95 percent confidence intervals are reported in brackets below the coefficient estimates. * denotes significance at the 95 percent level, and ** denotes significance at the 99 percent level. $\hat{\sigma}^2$ denotes the standard error of the regression – roughly interpreted as the average difference between the observed outcome and the prediction of the regression. All standard errors are clustered by journalist.

Table A.5 presents the results of an alternative specification designed to address this problem. In this analysis, we examine only the set of authors that eventually receive training. We then compare the quality of articles produced by authors who have already

Table A.4: List of Newspapers in Sample and Partisan Affiliation

Alasiri	Neutral
Changamoto	Pro-Opposition
Daily News	Pro-Government
Habari Leo	Pro-Government
Kulikoni	Pro-Opposition
Majira	Neutral
Mtanzania	Neutral
Mwana Halisi	Pro-Opposition
Mwanachi	Neutral
Nipashe	Neutral
Rai	Pro-Opposition
Rai Mwema	Neutral
Sauti ya watu Tanzani	Pro-Opposition
Sunday Citizen	Neutral
Sunday News	Pro-Government
Tanzania Daima	Pro-Opposition
The African	Neutral
The Citizen	Neutral
The East African	Neutral
The Guardian	Neutral
This Day	Pro-Opposition
Uhuru	Pro-Government
Uwazi	Neutral

Table A.5: Estimated Effect of Training on Article Quality – Trained Authors Only

	Baseline Model	Partisan Model
Training	-0.007 [-0.415,0.401]	0.068 [-0.500,0.636]
Pro Opposition Paper		0.067 [-0.729,0.864]
Pro Gov't Paper		0.137 [-0.317,0.592]
Pro Opposition Paper * Training		0.121 [-0.989,1.230]
Pro Gov't Paper * Training		-0.651* [-1.173,-0.129]
Time Fixed Effects	✓	✓
$\hat{\sigma}^2$	0.611	0.591

received training to that of authors who will receive training at some point in the future. So long as the method for selection into training remains consistent over time, these results are unlikely to be biased by selection effects.

The results reported in Table A.5 are largely unchanged from those reported in Section 3.4.1. The average effect of training is still statistically insignificant, and slightly negative. The most substantial change is in the estimated effect of training on articles appearing in pro-opposition papers, which is now estimated to be positive (though is not statistically significant). Training significantly *reduces* the quality of articles appearing in pro-government papers.

A.4 Public Expenditure Tracking

In this portion of the appendix, we discuss statistical details related to the analysis of the effect of PETS training on local government accounting. For full details of this analysis, see Section 3.4.2. Recall that this analysis relied on audits of local government accounts produced by the Controller and Auditor General (CAG). In Section 3.4.2, we present an analysis of the effect of PETS training on changes in the percentage of questioned expenditures and on the auditor opinions contained in these reports. Below, we report the details of the propensity score matching algorithm used to adjust for possible selection effects in assignment to PETS training. And we present the results from our analysis run on a matched dataset.

Matching Diagnostics

Recall that our analysis compares changes in auditor scores of local government accounts between districts that received PETS training and those that did not. This analysis, contained in Section 3.4.2, may be problematic because districts were not selected at random to receive PETS training and our outcome measures are constrained in the range of possible values they can assume. The former concern is relevant because trained districts may systematically differ from untrained districts. The second concern is relevant because it implies that the *level* of a given score can affect its rate of change. Districts that score extremely highly or extremely lowly on auditor reports can only change their scores in a single direction. Thus, the difference-in-differences analysis does not fully control for district specific factors that remain constant over time.

To help address these concerns, we turn to a propensity score matching technique. Matching serves to balance a dataset on key observable variables – so that like units are being compared. It may thus help to control for selection on observable covariates.¹⁰⁷ To conduct matching, we estimate the probability that a given district receives PETS training based on a broad range of data collected from pre-2007 CAG reports. We then weight each observation – using the genetic matching algorithm of Diamond and Sekhon (2006) – to ensure covariate balance in our sample. Districts that did not receive training that are insufficiently similar to trained districts are pruned from the dataset.

Figure A.7 presents a plot of propensity scores in our sample, and compares these be-

¹⁰⁷Unlike a Heckman selection procedure, matching does not control for selection on unobserved characteristics. However, when using a Heckman analysis, the analyst must make very strong assumptions regarding the selection process to derive her estimates.

tween trained (treated), untrained (control), and pruned (unmatched control) districts. The weights given to each observation are proportional to the size of the diamonds presented in the Figure.

Figure A.7: Distribution of Propensity Scores from Genetic Matching Algorithm

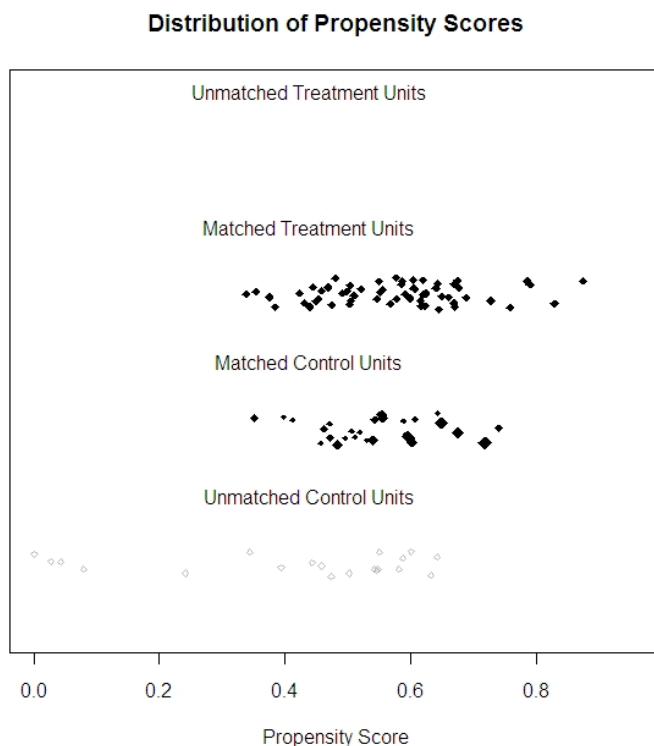
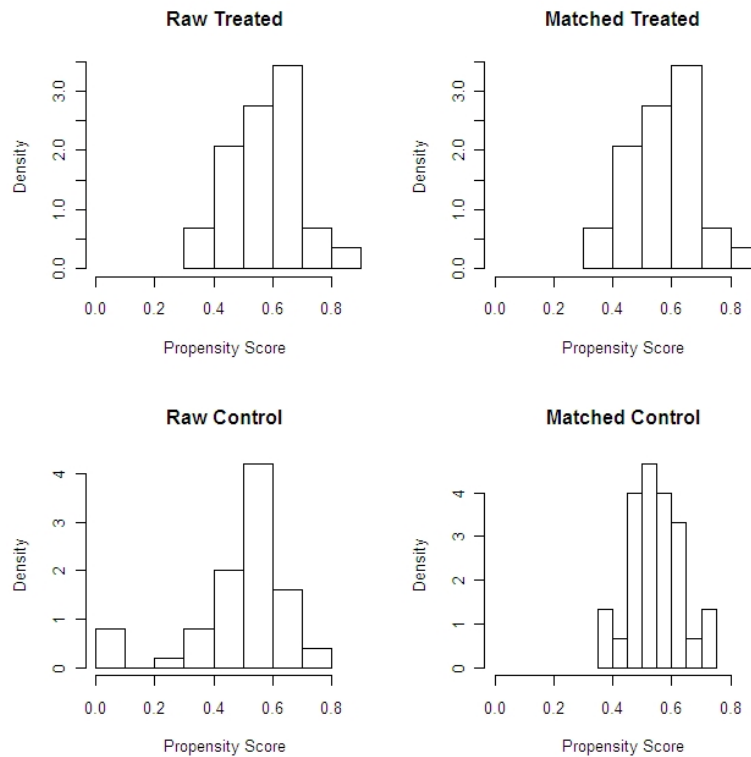


Figure A.8 presents histograms of the distribution of propensity scores before and after reweighting. In a perfectly balanced dataset, the distributions of propensity scores would be identical in trained (treated) and untrained (control) districts. As can be seen from the figure, this is certainly not the case prior to matching. The control group contains a number of districts with low propensity scores. After matching, the distributions of propensity scores are substantially more similar across the treated and control groups.

Matching Results

In this section, we present the results from regressions run on the matched dataset. In these regressions, we employ the same difference-in-differences estimator described in Section 3.4.2. Tables A.6 and A.7 report the results from these regressions.

Figure A.8: Histogram of Propensity Scores from Genetic Matching Algorithm



Our findings are largely unchanged. PETS training produces a small, but insignificant, increase in the percentage of questioned revenue. The magnitude on the coefficient in Model 1 is roughly twice that in the unmatched dataset. But it is still far from significant. PETS training is also associated with a small and insignificant decline in auditor opinions in both PACT and CGG districts – though CGG districts perform slightly better than PACT districts.

Qualitative Interviews

Finally, in Section 3.4.2, we report the results from a series of qualitative semi-structured interviews with PETS council chairs. PETS districts were selected by random sampling to be interviewed. And all interviews were administered by EcomResearch. Table A.8 lists the set of PETS council chairs interviewed.

Table A.6: Change in the Percent of Questioned Expenditures Associated with PETS Training – Matched Data

	Model 1	Model 2
PETS training	0.018 [-0.021,0.056]	0.015 [-0.025,0.054]
CGG training		0.017 [-0.032,0.066]
$\hat{\sigma}^2$	0.068	0.068

Table A.7: Change in Auditor Opinions Associated with PETS Training – Matched Data

	Model 1	Model 2
PETS training	-0.130 [-0.657,0.397]	-0.142 [-0.690,0.406]
CGG training		0.065 [-0.794,0.924]

Table A.8: List of PETS Councils Surveyed

Kibondo
Chamwino
Dodoma Urban
Iramba District
Kasulu
Mpanda
Sumbawanga Urban
Lindi Municipal Council
Rufiji
Moshi Rural
Muheza
Kilindi District Council
Lindi Rural District Council
Kongwa
Moshi Municipal Council

B PCCB Bureau Chief Questionnaire

Leonard Wantchekon (New York University); James R. Hollyer (New York University)

Interview outline for PCCB Bureau Chiefs.

Training provided by the U.S. Department of Justice.

1. Did bureau chiefs of the regional PCCB offices received the training provided by the U.S. Department of Justice between 2006 and 2008? (Yes, No)? _____

1. Je wakuu wa vituo vya mikoa vya PCCB walipata mafunza yaliyotolewa na idara ya usalama ya Marekani kati ya mwaka 2006 na 2008? Ndiyo/Hapana

a. If yes, can you please tell us the number of sessions attended?

a. Kama ndiyo, unaweza kututajia na mara ngapi ulihudhuria mafunzo?

b. If you do not remember the exact number of sessions attended. Can you provide us with an approximate number?

b. Kama hukumbuki idadi halisi tupe makadirio.

2. Can you please tell us about the content of the sessions attended and the type of training received?

2. Je unaweza kuelezea mambo mliyojifunza na aina ya mafunzo uliyohuduria?

3. In your opinion, would you consider the training received was useful in prosecuting corruption cases in the region of competence? How? _____

3. Kwa mapendekezo yako, unafikiri kuwa mafunzo uliyopata yalikuwa na manufaa katika kuendesha kesi za rushwa? Kwa vipi?

Training provided by the NYC Department of Investigations.

Mafunzo yaliyotolewa na idara ya NYC kitengo cha uchunguzi

4. Did members of the regional PCCB offices received any training provided by the NYC Department of Investigations during 2007? (Yes, No). _____

4. Je wafanyakazi wa TAKUKURU mkoani walipata mafunzo yaliyotolewa na idara ya NYC kitengo cha uchunguzi katika mwaka 2007? (Ndiyo/Hapana)

a. If yes, can you please tell us the number of sessions they attended?

a. Kama ndiyo, unaweza kutaja idadi ya mafunzo waliyohudhuria?

b. If you do not remember the exact number of sessions attended. Can you provide us with an approximate number?

b. Kama hukumbuki idadi halisi tupe makadirio.

5. Can you please tell us about the content of the sessions attended and the type of training received?

5. Je unaweza kuelezea mambo mliyojifunza na aina ya mafunzo uliyohudhuria?

6. In your opinion, would you consider the training received was useful in prosecuting corruption cases in the region of competence? How? _____

6. Kwa mapendekezo yako, unafikiri kuwa mafunzo uliyopata yalikuwa na manufaa katika kuendesha kesi za rushwa? Kwa vipi?

Training in general.

Mafunzo kwa ujumla

7. Did any of your **subordinates** in the regional PCCB office receive any training from, either the U.S. Department of Justice or NYC Department of Investigations? (Yes, No) _____

7. Je kuna mfanyakazi yeyote katika ofisi ya mkoa aliyepata mafunzo toka aidha kitengo cha haki U.S au idara ya uchunguzi N.Y? (Ndiyo/Hapana)

8. If for any training session your subordinates did **not** receive any training, was there any attempt to convey what they learnt to their subordinates? (Yes, No) _____

8. Kama kwa namna yoyote wafanyakazi hawakupata mafunzo, kulikuwa na jitihada zozote kuwafundisha mlichijifunza? (Ndiyo/Hapana)

9. If there was such attempt, could you describe **how** bureau chiefs conveyed what they learned to their subordinates? _____

9. Kama kulikuwa na juitihada hizo, unaweza kuelezea namna wakuu wa vituo walivyowafundisha wafanyakazi wao?

10. What kind of **training** activities continues to this day in the PCCB Regional Office? _____

10. Aina gani ya mafunzo yanaendelea mpaka leo katika ofisi za PCCB mikoani?

Post-Training PCCB performance.

Utendaji wa PCCB baada ya mafunzo

11. To the best of your knowledge, has the Regional Office of the PCCB prosecuted any corruption cases based on circumstantial evidence in the time after the training was received (2007-today)?

11. Kwa upeo wa uelewa wako ofisi za mikoa za PCCB zimeendesha mashitaka yoyote ya rushwa kwa ushahidi wa mazingira baada ya kupata mafunzo (2007-leo)?

12. Would you consider that the ability of the Regional Office of the PCCB to process cases based on circumstantial evidence has improved over time? (Yes, No) _____

12. Je unaona kuwa uwezo wa ofisi za mikoa za PCCB kufanya uchunguzi kwa kuzingatia ushahidi wa mazingira umeongezeka? (Ndiyo/Hapana)

a. If yes, would you attribute any improvement in the ability to process cases to the training received by either the U.S. Department of Justice or NYC Department of Investigations?

a. Kama ndiyo unaweza kuhusianisha kuongezeka huko na mafunzo yaliyotolewa na idara ya uchunguzi NYC?

b. Specifically, how do you apply in your everyday work what was learnt during the training? Can you provide a concrete example? _____

b. Je unatumiaje ujuzi uliopata tokana na mafunzo katika kazi zako za kila siku? Unaweza kutoa mifano?

13. What would you consider are the limitations of the training received by either the US Department of Justice or the NYC Department of Investigations? _____

13. Nini unaona ni vipingamizi wa mafunzo uliyopata?

a. How would you improve the implementation of these training programs? _____

a. Unawezaje kuboresha utekelezaji wa mafunzo haya?

14. When prosecuting cases based on circumstantial evidence, which are the most common obstacles you encounter? (e.g. judges are reluctant to hear such cases, prosecutors are reluctant to take these cases, among others). Please be as specific as possible. _____

14. Mnapoendesha mashtaka za ushahidi wa mazingira ni vipingamizi gani vikubwa unavyokutana navyo? (mf. majaji ni wazito kusikiliza kesi hizo, waendesha mashtaka hawapendi kuendesha kesi hizo, mengineyo)

15. In your opinion, which are the areas where the PCCB office finds more evidence of corruption, misallocation of funds, among others? For example, government procurement cases in construction contracts versus purchasing government supplies. _____

15. Kwa mawazo yako, ni maeneo gani ambayo PCCB wanapata ushahidi wa rushwa, matumizi mabaya ya fedha, mengineyo? Mfano, manunuzi ya serikali katika mimkataba ya ujenzi?

16. Do you consider that the passage of the Prevention and Combating of Corruption Act (2007) has improved the ability of the PCCB to prosecute corruption cases? _____

16. Unafikiria kupitishwa kwa sheria ya PCCCB ya mwaka 2007 kumeongeza uwezo wa PCCB kubaini mashitaka ya kesi za rushwa?

17. If yes, how?

17. Kama ndiyo kwa vipi?

18. In your opinion, has the new provision of the law expanding the number and types of charges that can be prosecuted improved your ability to rule in a broader number of cases?

18. Kwa maoni yako, sheria mpya inayoongeza idadi ya makosa yanayoweza kushtakiwa imeongeza uwezo wa maamuzi katika kesi nyingi zaidi?

19. Would you consider that your subordinates or members of the staff are familiar with provisions contained in the Prevention and Combating of Corruption act of 2007? (Yes, No)

19. Je unafikiri kuwa wafanyakazi wa ofisi yako wana uelewa wa yaliyomo katika sheria ya PCCB ya mwaka 2007?

a. How has this knowledge affected investigations and prosecutions? Can you provide a specific example?

a. Je uelewa huu umeleta mabadiliko gani katika uchunguzi? Unaweza kutoa mifano halisi?

b. Has a lack of knowledge of the specifics of this law hindered your ability to investigate and prosecute these cases? _____

b. Je kutokuwa na uelewa halisi wa sheria hii unazuia uwezo wako wa uchunguzi na kufanya mashaka?

C Questionnaire for PETS Council Chairs

Leonard Wantchekon (New York University); James R. Hollyer (New York University)

Interview outline for PETS Council chairs

Meetings per month (Frequency)

1 How many meetings does the PETS council hold per month? If you do not know the exact number, could you give us an approximation?

Size of the PETS Council (Number of members).

2 What is the number of members of the PETS council? If you do not know the exact number, could you give us an approximation?

Attendance (Perception).

3 Can you provide us with the percentage of members from the council **regularly** attending PETS council meetings? (by regularly we mean more than 50% of the time).

4 Do you consider non-attendance of members (absenteeism) a problem affecting PETS council's performance?

Areas of Budgeting Concentration.

5. What would you consider the main areas on which the PETS council spends most of the time of their meetings? (Few words). _____

6. How does the PETS council decide which budgeting areas to concentrate on?

7. How do they reach this decision? (e. g. by majority vote, unanimity, among others.) _____

8. Has the PETS council found any discrepancies (e.g. misallocation of funds, unclear assignment of resources, among others) since it began operations?

a) Yes → Go to Question 8a, 8b and 8c.

b) No → Go to Question 9.

8a. *Approximately* how many discrepancies have been found by the PETS council since it began operations?

8b. Were these discrepancies publicized in the local media or on the PETS notice boards?

8c. Which **type** of discrepancies were the most common, that is, in which budgeting areas they were more frequent?

Referring to the education level of members of the PETS Council:

9. Which of the following is your highest degree of education? (To council chair)

- a) Primary School
- b) Secondary
- c) High School
- d) Technical Career
- e) Undergraduate degree.
- f) Graduate degree.

g) Other. _____

10 Which of the following do you consider is the **average** level of education of **the majority** of the members in the Board?

a) Primary School

b) Secondary

c) High School

d) Technical Career

e) Undergraduate degree.

f) Graduate degree.

g) Other _____

Referring to the budgeting experience of the members of the PETS Council, that is, whether members of the Board held any accounting or financial position previous to their engagement in the PETS Council:

11 What is the past budgeting experience of the most experienced member of the council? For example, previous experience as an accountant, in business administration or as a financial planner for private or non-profit organization. _____

Referring to the level of engagement of PETS council with local villagers:

12. Do district level PETS councils regularly consider the needs expressed by local villagers?

a) Yes, if they do so in their everyday work → Go to Question 12a, 12b.

b) No, if they do not. → Go to Question 13.

12a. If yes, how **does** the PETS council consult with the villagers? For example, local villagers establish verbal contact with members of the council; local villagers present a written petition to the council; members of the board attend village meetings, among others. _____

12b. If yes, how **frequently** does the PETS council consult with the villagers?

Publicity of the PETS council activities:

13. Do you consider that PETS activities are properly publicized in the streets, local newspapers or media?

a) Yes

b) No

Procedural rules governing PETS council meetings:

14. Are PETS council meetings open to the public?

a) Yes

b) No

15. How are tasks assigned to council members?

16. Is a given task assigned to more than one member, or each task is executed by an individual?

17. Is there any written document of the procedures/bylaws by which the council is governed? If so, would it be possible to obtain a copy?

Perceived effect of PETS councils:

18. Do you consider that there has been a change in the local government behavior as a result of PETS council budgeting surveillance? If so, please explain briefly.

19. Has there been a change in the level of civic engagement as a result of PETS activities? That is, are people more willing to demand certain action from the local government? If so, can you provide a specific example?
